

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 23:30:20 ; Search time 7544.92 Seconds
(without alignments)
11245.517 Million cell updates/sec

Title: US-10-005-337A-2

Perfect score: 2074

Sequence: 1 ctgcagcaagttacttaatg.....acaagactcttcagccaac 2074

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 20454813386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rod:*

36: em_htg_mam:*

37: em_htg_vrt:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2071.6	99.9	2074	6	AX468604	AX468604 Sequence
2	2071.6	99.9	2074	9	AF131884	AF131884 Homo sapi
C 3	1395	67.3	158357	9	AL365434	AL365434 Human DNA
C 4	1395	67.3	160350	2	AC074094	AC074094 Homo sapi
C 5	664.8	32.1	50111	9	AL590622	AL590622 Human DNA
6	367.8	17.7	2358	6	AX468603	AX468603 Sequence
7	367.8	17.7	142902	2	AC119234	AC119234 Mus muscu
8	365	17.6	229640	2	AC105469	AC105469 Rattus no
9	365	17.6	238344	2	AC097115	AC097115 Rattus no
10	340.8	16.4	723	10	AF478692	AF478692 Mus muscu
11	228	11.0	1901	6	AX322775	AX322775 Sequence
12	228	11.0	1901	6	BD094076	BD094076 Shear str
13	228	11.0	1901	9	HSRNACINP	X83703 H.sapiens m
14	228	11.0	1901	11	G28603	G28603 human STS S
15	154.4	7.4	1940	4	AF131883	AF131883 Oryctolag
16	144.4	7.0	1889	6	AX281749	AX281749 Sequence
C 17	133.6	6.4	838	11	BV062893	BV062893 S212B6025
C 18	102.2	4.9	210651	2	AC026715	AC026715 Homo sapi
C 19	101	4.9	160901	2	AC120060	AC120060 Rattus no
C 20	101	4.9	235419	2	AC095979	AC095979 Rattus no
C 21	100.6	4.9	38476	9	U73024	U73024 Homo sapien
C 22	99	4.8	116803	2	AC097718	AC097718 Homo sapi
C 23	99	4.8	157665	9	AC104695	AC104695 Homo sapi
C 24	98.8	4.8	87402	9	HSJ437M21	AL049758 Human DNA
C 25	98.4	4.7	186870	9	AC026471	AC026471 Homo sapi
C 26	98.4	4.7	227074	2	AC122863	AC122863 Mus muscu
C 27	98.2	4.7	51529	9	AL591291	AL591291 Human DNA
C 28	98.2	4.7	148259	10	AC104099	AC104099 Mus muscu
C 29	98	4.7	172260	9	AC068946	AC068946 Homo sapi
C 30	97.8	4.7	152696	2	AC129985	AC129985 Homo sapi
C 31	97.8	4.7	153053	9	AC090711	AC090711 Homo sapi
C 32	97.8	4.7	163433	2	AC079993	AC079993 Homo sapi
C 33	97.8	4.7	179040	2	AC022732	AC022732 Homo sapi
C 34	97.6	4.7	51370	2	AC084305	AC084305 Homo sapi
C 35	97.6	4.7	218309	2	AC137376	AC137376 Rattus no
C 36	97.6	4.7	249682	2	AC096394	AC096394 Rattus no
C 37	97.4	4.7	5404	9	HSN804865	AL833552 Homo sapi
C 38	97.4	4.7	34917	9	AL390211	AL390211 Human DNA
C 39	97.2	4.7	34488	6	BD093713	BD093713 Human lp3
C 40	97.2	4.7	90289	2	AP003164	AP003164 Homo sapi
C 41	97.2	4.7	121162	6	AX039602	AX039602 Sequence
C 42	97.2	4.7	171612	9	AL358013	AL358013 Human DNA
C 43	97	4.7	124321	9	HS417G15	AL009174 Human DNA
C 44	97	4.7	163338	9	AL162426	AL162426 Human DNA
C 45	96.8	4.7	142952	2	AC040909	AC040909 Homo sapi

ALIGNMENTS

RESULT 1
AX468604
LOCUS
DEFINITION Sequence 2 from Patent WO0246220.
ACCESSION AX468604
VERSION AX468604.1 GI:21901403
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Schwartz,B., Branellec,D. and Chien,K.
TITLE Sequences upstream of the carp gene, vectors containing them and
uses thereof

JOURNAL Patent: WO 0246220-A 2 13-JUN-2002;
Aventis Pharma S.A. (FR); The Regents of The University of
California at San Diego (US); Benoit, Patrick (FR)
FEATURES Location/Qualifiers
source 1..2074
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 612 a 469 c 416 g 572 t 5 others
ORIGIN
Query Match 99.9%; Score 2071.6; DB 6; Length 2074;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2074; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CTCGACGAGTTACTTAATGTTTTTGGCTCAGCATCTCTCTGTAAATGAGAGCATTA 60
Db 1 CTGCGCAAGTTACTTAATGTTTTTGGCTCAGCATCTCTCTGTAAATGAGAGCATTA 60
Qy 61 GTCTTGTCTCAACTTCGAGGCGATGGACAGCTCTGGGATTTCAATCCAGAGCCTTAA 120
Db 61 GTCTTGTCTCAACTTCGAGGCGATGGACAGCTCTGGGATTTCAATCCAGAGCCTTAA 120
Qy 121 CATCCACAGTCTCTCCCAACACCTTCTCCTCTAATACCTCCCTCAGTTTGGGTGAG 180
Db 121 CATCCACAGTCTCTCCCAACACCTTCTCCTCTAATACCTCCCTCAGTTTGGGTGAG 180
Qy 181 GCCTGGAAACAAAAGGCATACGAAATGGTAAAGAGTGTCCATGACTTCTCGACTTA 240
Db 181 GCCTGGAAACAAAAGGCATACGAAATGGTAAAGAGTGTCCATGACTTCTCGACTTA 240
Qy 241 GATGAAGAGACCAATGAAATAGTAATGACTCTGTGTCGAGGAGACATATACTAAA 300
Db 241 GATGAAGAGACCAATGAAATAGTAATGACTCTGTGTCGAGGAGACATATACTAAA 300
Qy 301 ATAGGAGCTATACAAAGAGATAGCATGAGTCTGTGCAAGAAATGACACACAAATTTGT 360
Db 301 ATAGGAGCTATACAAAGAGATAGCATGAGTCTGTGCAAGAAATGACACACAAATTTGT 360
Qy 361 GAAACATTCATATATTAATAAATAAATAAATAAAGAGAAAGGAAATTAATAAG 420
Db 361 GAAACATTCATATATTAATAAATAAATAAATAAAGAGAAAGGAAATTAATAAG 420
Qy 421 AAAATAGTAGTGTGTCTCATCTCAAGAAAGCCAGGAGATTTCTTTATTTACCC 480
Db 421 AAAATAGTAGTGTGTCTCATCTCAAGAAAGCCAGGAGATTTCTTTATTTACCC 480
Qy 481 CCTTTAAGATAGAAATATTAGGAGACCGGAACATATGATCAGGAGGTACTGGGAGGTCC 540
Db 481 CCTTTAAGATAGAAATATTAGGAGACCGGAACATATGATCAGGAGGTACTGGGAGGTCC 540
Qy 541 CTCTTTGTCAATGTTTTGTCTTGGGTTGGGAGTTCGATGCTTTCTCAAGATTTTCAAGAAC 600
Db 541 CTCTTTGTCAATGTTTTGTCTTGGGTTGGGAGTTCGATGCTTTCTCAAGATTTTCAAGAAC 600
Qy 601 ACCATCCACTGACTGAGCATTTCAAGGGCAAGAGGAGATGGCAGCCACATTTGTGATT 660
Db 601 ACCATCCACTGACTGAGCATTTCAAGGGCAAGAGGAGATGGCAGCCACATTTGTGATT 660
Qy 661 GGGTGAAGTTGGGAGAGAAATAGACACAAAGGTCAAAACATACTTCTTAATTAACACTT 720
Db 661 GGGTGAAGTTGGGAGAGAAATAGACACAAAGGTCAAAACATACTTCTTAATTAACACTT 720
Qy 721 CCTTCCATTCACAAATTCCTTCTCCCAATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 780
Db 721 CCTTCCATTCACAAATTCCTTCTCCCAATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 780
Qy 781 AGTTTTTCTGAAACTATAAAATACCCCAAGTATGTTTACATAATTTACACCTCAAGA 840
Db 781 AGTTTTTCTGAAACTATAAAATACCCCAAGTATGTTTACATAATTTACACCTCAAGA 840
Qy 841 TTAGAACCAAGAAATAGAGACCTTTTCAACCTTCCGGAAGCAAGTGCATTTATCCCTCC 900
Db 841 TTAGAACCAAGAAATAGAGACCTTTTCAACCTTCCGGAAGCAAGTGCATTTATCCCTCC 900

Db 841 TTAGAACCAAGAAATAGAGACCTTTTCAACCTTCCGGAAGCAAGTGCATTTATCCCTCC 900
Qy 901 AGCCAGTGTCTCAAAATCTTGTATGATCAGATCATCTGGTGCTTTTAAATTCAGATG 960
Db 901 AGCCAGTGTCTCAAAATCTTGTATGATCAGATCATCTGGTGCTTTTAAATTCAGATG 960
Qy 961 ATTCTTACGAGTTTACCATAAACTCAAGATTTCCCTGGAGTGGGCGCAGGATCTGTA 1020
Db 961 ATTCTTACGAGTTTACCATAAACTCAAGATTTCCCTGGAGTGGGCGCAGGATCTGTA 1020
Qy 1021 TTTCTGACAAGCTCCCAACAGGTGATTTCTTCCCAACAGCATTTGAGAACTTCAGCTCAA 1080
Db 1021 TTTCTGACAAGCTCCCAACAGGTGATTTCTTCCCAACAGCATTTGAGAACTTCAGCTCAA 1080
Qy 1081 TGACCTAATCAGAGTCTCGCCATTTGCTAATAATCTGTCTCATTTTTTBTCTATATATA 1140
Db 1081 TGACCTAATCAGAGTCTCGCCATTTGCTAATAATCTGTCTCATTTTTTBTCTATATATA 1140
Qy 1141 TAGTATTTGTTGGTAGAGATGGGATTTTGCATGTTGCCAGGCTAGTATTGAACCTCTAA 1200
Db 1141 TAGTATTTGTTGGTAGAGATGGGATTTTGCATGTTGCCAGGCTAGTATTGAACCTCTAA 1200
Qy 1201 GCTAAGCAATCTTCTGTCTCTGCCCTCCCAAAATGTTGGGATTAACAGGCTTAAGCCACTG 1260
Db 1201 GCTAAGCAATCTTCTGTCTCTGCCCTCCCAAAATGTTGGGATTAACAGGCTTAAGCCACTG 1260
Qy 1261 CACCCGGCTGATAGCTGGTTTCAATTTACTTCTATTTCTTGACCACTCTGATCCATTTGAA 1320
Db 1261 CACCCGGCTGATAGCTGGTTTCAATTTACTTCTATTTCTTGACCACTCTGATCCATTTGAA 1320
Qy 1321 GTAAAAATGCTCCAAATTTATGCTGTTTGTAGAACACGGTAAAGCATGTCAATGCTAAATG 1380
Db 1321 GTAAAAATGCTCCAAATTTATGCTGTTTGTAGAACACGGTAAAGCATGTCAATGCTAAATG 1380
Qy 1381 GCCAGTGCATCATATAAAGAAAGTGCATTTACTGAATGCTTTCAATGCTTTTATAATGATG 1440
Db 1381 GCCAGTGCATCATATAAAGAAAGTGCATTTACTGAATGCTTTCAATGCTTTTATAATGATG 1440
Qy 1441 GTAAGTGGATGTCATGGGGCTATTTAGCCGACATCATCTCCAAAGAAATTCCAAACA 1500
Db 1441 GTAAGTGGATGTCATGGGGCTATTTAGCCGACATCATCTCCAAAGAAATTCCAAACA 1500
Qy 1501 GATATAGACAAGTGCCTTTAGGGCCAGATCCCTTCCCTCAGGCTGTTTACCCAGGAA 1560
Db 1501 GATATAGACAAGTGCCTTTAGGGCCAGATCCCTTCCCTCAGGCTGTTTACCCAGGAA 1560
Qy 1561 TAGGATGTCTCTGGGAACAAGTTTCCCTTAAGTGAAGTGTGATAAGTCTGCTTATCAGAA 1620
Db 1561 TAGGATGTCTCTGGGAACAAGTTTCCCTTAAGTGAAGTGTGATAAGTCTGCTTATCAGAA 1620
Qy 1621 GATATTACTGGGGGTGTGATATGAGGGCATCTACATTTTCTTGATAGGTAGTCATATGA 1680
Db 1621 GATATTACTGGGGGTGTGATATGAGGGCATCTACATTTTCTTGATAGGTAGTCATATGA 1680
Qy 1681 AAGCTCACAAGAAAGAAAGGCGAGTGTGTGCAATGTCTCAACAGACAGCTGTCCCT 1740
Db 1681 AAGCTCACAAGAAAGAAAGGCGAGTGTGTGCAATGTCTCAACAGACAGCTGTCCCT 1740
Qy 1741 GACTCTTGACAAATAGGATGACTTGCATTTGCTGAGCGATGTGATCACCACCAAGGAATG 1800
Db 1741 GACTCTTGACAAATAGGATGACTTGCATTTGCTGAGCGATGTGATCACCACCAAGGAATG 1800
Qy 1801 GCCTCTCACATTTCTTCTGTATTCATATTTTCCAGCAGGTTAGCTTGTCTTCCCTCCC 1860
Db 1801 GCCTCTCACATTTCTTCTGTATTCATATTTTCCAGCAGGTTAGCTTGTCTTCCCTCCC 1860
Qy 1861 TCTTCAAGTTTCCAGACACTGAGTCTGGAATGAAATTCACCTGCTCTGAGTTGGCTCC 1920
Db 1861 TCTTCAAGTTTCCAGACACTGAGTCTGGAATGAAATTCACCTGCTCTGAGTTGGCTCC 1920
Qy 1921 TAATGGGGCGGAGTGTACTTTCGGTTCCCAAGTTTGAAGATTTACTCACCCGGCCCCA 1980
Db 1921 TAATGGGGCGGAGTGTACTTTCGGTTCCCAAGTTTGAAGATTTACTCACCCGGCCCCA 1980

```

QY 1981 GCTATATAAGCTGACCGGTGTGAGGGGCCAGCAGGGCCAACTCCAGGGATTCTTCCA 2040
DB 1981 GCTATATAAGCTGACCGGTGTGAGGGGCCAGCAGGGCCAACTCCAGGGATTCTTCCA 2040

QY 2041 CGACAGAAAACATACAGACTCTTTCAGCCAAC 2074
DB 2041 CGACAGAAAACATACAGACTCTTTCAGCCAAC 2074

RESULT 2
AF131884
LOCUS AF131884 2074 bp DNA linear PRI 09-FEB-2000
DEFINITION Homo sapiens cardiovascular-specific cardiac ankyrin repeat protein
(CVAP) gene, 5'-flanking region and exon 1.
ACCESSION AF131884
VERSION AF131884.1 GI:6940841
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2074)
AUTHORS Aihara,Y., Kurabayashi,M., Tanaka,T., Sekiguchi,K., Tomaru,K.,
Kanai,H., Takeda,S. and Nagai,R.
TITLE Human CVARP 5'-flanking region
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2074)
AUTHORS Aihara,Y.
DIRECT SUBMISSION
SUBMITTED (26-FEB-1999) Second Department of Internal Medicine,
Gunma University School of Medicine, 3-39-15 Showa-machi, Maebashi
371-8511, Japan
FEATURES
source 1..2074
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
misc_feature 1..1832
/note="5'-flanking region"
gene 1833..>2074
/gene="CVARP"
mRNA 1833..>2074
/gene="CVARP"
/product="cardiovascular-specific cardiac ankyrin repeat
protein"
/note="CARP"
exon 1833..2074
/gene="CVARP"
/number=1
BASE COUNT 612 a 469 c 416 g 572 t 5 others
ORIGIN
Query Match 99.9%; Score 2071.6; DB 9; Length 2074;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2074; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCAGCAAGTACTTAATGTTTTTGGCTCAGCATCCTCTCTGTTAAATGAGAGCATTA 60
DB 1 CTGCAGCAAGTACTTAATGTTTTTGGCTCAGCATCCTCTCTGTTAAATGAGAGCATTA 60

QY 61 GTCTTGCTCCAACTTCGAGGGCATGACAGCTCTGGGATTTTATATCCAAAGACCCCTTAA 120
DB 61 GTCTTGCTCCAACTTCGAGGGCATGACAGCTCTGGGATTTTATATCCAAAGACCCCTTAA 120

QY 121 CATCCACAGTCTTCCCCCAACACTTCTCTCTTAATACCTCCCTCAGTTTGGGTGAG 180
DB 121 CATCCACAGTCTTCCCCCAACACTTCTCTCTTAATACCTCCCTCAGTTTGGGTGAG 180

QY 181 GCCTGAAACAAAAGGCATACGAAATGGTAGAAAAAGTGTCCATGACTACTTCTGACTTA 240
DB 181 GCCTGAAACAAAAGGCATACGAAATGGTAGAAAAAGTGTCCATGACTACTTCTGACTTA 240

```

```

QY 241 GATGAAGAGACCAATGAAATAATAGTAACTCTGTTTGTCTTTCAGCAGGACATATACTAA 300
DB 241 GATGAAGAGACCAATGAAATAATAGTAACTCTGTTTGTCTTTCAGCAGGACATATACTAA 300

QY 301 ATAGGAGCTATACAAGAGATTAGCATGGACTCTGTGCAAGAAATGACACACAAATTTGT 360
DB 301 ATAGGAGCTATACAAGAGATTAGCATGGACTCTGTGCAAGAAATGACACACAAATTTGT 360

QY 361 GAAACATTCATATATTTAAAAATAATAAATAAAGAGAAAGGAAAAAATTTAAAAAG 420
DB 361 GAAACATTCATATATTTAAAAATAATAAATAAAGAGAGAAAGGAAAAAATTTAAAAAG 420

QY 421 AAAATAGTATAGCTGTGTTCATCTCAAGAAAAAGCCAGGAGATTTCCTTTATTTACCC 480
DB 421 AAAATAGTATAGCTGTGTTCATCTCAAGAAAAAGCCAGGAGATTTCCTTTATTTACCC 480

QY 481 CCTTTAAGATCAATATTTAGGAGACCGGAACATATGATACAGAGGTACTGGGAGGTCC 540
DB 481 CCTTTAAGATCAATATTTAGGAGACCGGAACATATGATACAGAGGTACTGGGAGGTCC 540

QY 541 CTCCTTTGTCATATGTTTGTCTTGGGTGGGAGTGCATGCTTCTCAAAAGTTTTCAGAAAC 600
DB 541 CTCCTTTGTCATATGTTTGTCTTGGGTGGGAGTGCATGCTTCTCAAAAGTTTTCAGAAAC 600

QY 601 ACCATCCACTGACTGAGCAATTCAGGGGCAAGAGAGAAATGCGAGCCACATTTGTTGATT 660
DB 601 ACCATCCACTGACTGAGCAATTCAGGGGCAAGAGAGAAATGCGAGCCACATTTGTTGATT 660

QY 661 GGGTGAGTTGGGGAGAAATAGACACACAAAGGTCAACATACTTCCTTAATTTAACACTT 720
DB 661 GGGTGAGTTGGGGAGAAATAGACACACAAAGGTCAACATACTTCCTTAATTTAACACTT 720

QY 721 CCTCCATTCACAAATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 780
DB 721 CCTCCATTCACAAATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 780

QY 781 AGTTTTTCTGAAACTATATAAAATAACCCCACTAGTGTGTTTACATAATTTACACCTCAAAGA 840
DB 781 AGTTTTTCTGAAACTATATAAAATAACCCCACTAGTGTGTTTACATAATTTACACCTCAAAGA 840

QY 841 TTAGAAACCGAGAAATAGAGACCTTTTCAACCTTTCGGGAAGCAAGTGCATATCCCTCC 900
DB 841 TTAGAAACCGAGAAATAGAGACCTTTTCAACCTTTCGGGAAGCAAGTGCATATCCCTCC 900

QY 901 ACCCAGTGTCTCAAACTCTTGATGATCAGAAATCATCTGGGTGCTTTKAAATTTCAAGATG 960
DB 901 ACCCAGTGTCTCAAACTCTTGATGATCAGAAATCATCTGGGTGCTTTKAAATTTCAAGATG 960

QY 961 ATTCTAGAGTTACCATAAAATCAACTCAGAAATTCCTGGAGTGGGGCCAGGATCTGTA 1020
DB 961 ATTCTAGAGTTACCATAAAATCAACTCAGAAATTCCTGGAGTGGGGCCAGGATCTGTA 1020

QY 1021 TTTCTGACAGCTCCCAAGGTGATTCTTTTCCCAACAGCATTTTGAGAACTTCAGCTCAA 1080
DB 1021 TTTCTGACAGCTCCCAAGGTGATTCTTTTCCCAACAGCATTTTGAGAACTTCAGCTCAA 1080

QY 1081 TCACCTTAATCAGAGTCTGCCATTCCTATATCTCTGCTCTCATTTTBTCTATATATATA 1140
DB 1081 TCACCTTAATCAGAGTCTGCCATTCCTATATCTCTGCTCTCATTTTBTCTATATATATA 1140

QY 1141 TAGTATTTTGTGTAGAGATGGGATTTTGCCATGTTGCCAGGCTAGTATTTGAACCTCTAA 1200
DB 1141 TAGTATTTTGTGTAGAGATGGGATTTTGCCATGTTGCCAGGCTAGTATTTGAACCTCTAA 1200

QY 1201 GCTAAGCAATCTTCTGCTCTGCTCCCAAAATGTTGGGATTTACAGGTGTGAAGCCACTG 1260
DB 1201 GCTAAGCAATCTTCTGCTCTGCTCCCAAAATGTTGGGATTTACAGGTGTGAAGCCACTG 1260

QY 1261 CACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCTCATTTTGA 1320
DB 1261 CACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACTCTGATCTCATTTTGA 1320

QY 1321 GTAAAAATGCTCCCAATTATTATGCTGTTTGTAGAACACGGTAAAGCATGTCTATGTGCTAA 1380

```

```

Db 1321 GTAAATGCTCAATATTATCTGTTTGTAGAACGCGTAAGCATGTCTATGTGCTAATG 1380
Qy 1381 GCAGTGACATCATATAAGAAAAGTGCATTAATCTGAATGCTTTCAATGTCTTATAATGATG 1440
Db 1381 GCAGTGACATCATATAAGAAAAGTGCATTAATCTGAATGCTTTCAATGTCTTATAATGATG 1440
Qy 1441 GTAAAGTGGCATCTCATGGGCGCTATTAGCCAGACATCACTCCAAAGAAATTCGAACA 1500
Db 1441 GTAAAGTGGCATCTCATGGGCGCTATTAGCCAGACATCACTCCAAAGAAATTCGAACA 1500
Qy 1501 GATATAGACAAGTGCCTTTAGGGCCCGAGATCCCTTCCCTCAGCTGTTTACCCAGGGA 1560
Db 1501 GATATAGACAAGTGCCTTTAGGGCCCGAGATCCCTTCCCTCAGCTGTTTACCCAGGGA 1560
Qy 1561 TAGATGTCCTGGGCAAGTTTCCCTTAAGTGAAGTGTGATGAAGTCTGCTTATCAGAAA 1620
Db 1561 TAGATGTCCTGGGCAAGTTTCCCTTAAGTGAAGTGTGATGAAGTCTGCTTATCAGAAA 1620
Qy 1621 GATATTACTGGGGGTGTGATATAGGGCATCTACATTTTCTTGATAGGTAGTCATATGA 1680
Db 1621 GATATTACTGGGGGTGTGATATAGGGCATCTACATTTTCTTGATAGGTAGTCATATGA 1680
Qy 1681 AAGCTGACAAAGAAAAGGCGAGTGTGTGCAATGTCAACAGACAGCTGTGCCCT 1740
Db 1681 AAGCTGACAAAGAAAAGGCGAGTGTGTGCAATGTCAACAGACAGCTGTGCCCT 1740
Qy 1741 GACTCTTGACAAATAGGATGACTTGCTGAGCGATGTGATCACCACCAAGGAATG 1800
Db 1741 GACTCTTGACAAATAGGATGACTTGCTGAGCGATGTGATCACCACCAAGGAATG 1800
Qy 1801 GCCTCTCACATTTCTCTCTGATTCATATTCAGCAGGTTAGCTTGCTCCCTCC 1860
Db 1801 GCCTCTCACATTTCTCTCTGATTCATATTCAGCAGGTTAGCTTGCTCCCTCC 1860
Qy 1861 TCTTCAGCTTCCAGACACTGAGTCTGGAATGAAATTCACCTGCCTCTGAGTTGGCTCC 1920
Db 1861 TCTTCAGCTTCCAGACACTGAGTCTGGAATGAAATTCACCTGCCTCTGAGTTGGCTCC 1920
Qy 1921 TAATGGGGCGGAGTGTTACTTCGTTCCAGGTTCCAGAGATTATCTCACCCGGCCCCA 1980
Db 1921 TAATGGGGCGGAGTGTTACTTCGTTCCAGGTTCCAGAGATTATCTCACCCGGCCCCA 1980
Qy 1981 GCTATATAAGCTACCGGTGTGAGGGGCCAGCAGGCCAACTCCAGGATTCCTTCCA 2040
Db 1981 GCTATATAAGCTACCGGTGTGAGGGGCCAGCAGGCCAACTCCAGGATTCCTTCCA 2040
Qy 2041 CGACAGAAAACATACAAGACTCCTTCAGGCCAAC 2074
Db 2041 CGACAGAAAACATACAAGACTCCTTCAGGCCAAC 2074

RESULT 3
AL365434/c 158357 bp DNA linear PRI 21-JUN-2002
LOCUS Human DNA sequence from clone RP11-236B18 on chromosome 10,
DEFINITION complete sequence.
ACCESSION AL365434
VERSION AL365434.13 GI:21540024
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 158357)
JOURNAL Direct Submission
Submitted (21-JUN-2002) Wellcome Trust Sanger Institute, Hinxton,
Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
On Jun 23, 2002 this sequence version replaced gi:12191663.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations

```

together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases:

Em: EMBL; Sw: SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP database can be found at

http://www.sanger.ac.uk/Projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at

http://www.sanger.ac.uk/HGP/Chr10

RP11-236B18 is from the library RP11-11.1 constructed by the group of Pieter de Jong. For further details see

http://www.chori.org/bacpac/home.htm

VECTOR: pBACE3.6

----- Genome Center

Center: Wellcome Trust Sanger Institute

Center code: SC

Web site: http://www.sanger.ac.uk

Contact: humquery@sanger.ac.uk

This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

FEATURES

Location/Qualifiers

source

```

1..158357
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-236B18"
/clone_lib="RP11-11.1"

```

```

BASE COUNT 46958 a 32356 c 32839 g 46204 t
ORIGIN

```

```

Query Match 67.3%; Score 1395; DB 9; Length 158357;
Best Local Similarity 98.4%; Pred. No. 7.2e-307;
Matches 1479; Conservative 4; Mismatches 10; Indels 10; Gaps 7;

Qy 1 CTGCAGCAAGTTACTTAATGTTTCTTTGGCTCAGCATCCTCTCTGTAATAATGAGAGCATT 59
Db 1502 CTGCAGCAAGTTACTTAATGTTTCTTTGGCTCAGCATCCTCTCTGTAATAATGAGAGCATT 1443

Qy 60 AGTCTTGCTCCAACTTCGAGGGCATGGACAGCTCTGGGATTTTCATATCCAAAGACCTTAA 119
Db 1442 AGTCTTGCTCCAACTTCGAGGGCATGGACAGCTCTGGGATTTTCATATCCAAAGACCTTAA 1383

Qy 120 ACATCCCAAGTCTCTTCCCAAGCACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 179
Db 1382 ACATCCCAAGTCTCTTCCCAAGCACTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1323

Qy 180 GGCCTGGAAACAAAAGGCATACGAAATGGTAGAAAAGTGTCCATGACTACTTCTGACTT 239
Db 1322 GGCCTGGAAACAAAAGGCATACGAAATGGTAGAAAAGTGTCCATGACTACTTCTGACTT 1263

Qy 240 AGATGAAGAGACCAATGAAAATAGTAACTGCTCTGTTTCTTCAGCAGACATATATACTAA 299
Db 1262 AGATGAAGAGACCAATGAAAATAGTAACTGCTCTGTTTCTTCAGCAGACATATATACTAA 1203

Qy 300 AATAGGAGCTATACAAAGAAAGATTAGCATGGACTCTGTGCAAGAATGACACACAAATTTG 359
Db 1202 AATAGGAGCTATACAAAGAAAGATTAGCATGGACTCTGTGCAAGAATGACACACAAATTTG 1143

Qy 360 TGAACATCTCCATATATTAATAATAATAATAATAATAATAAGAGAAAAGAAAAATTAATAA 419
Db 1142 TGAACATCTCCATATATTAATAATAATAATAATAATAAGAGAAAAGAAAAATTAATAA 1083

```


Db 113800 CCTAAGCTAAGCAATCTCTCTGCTCTGCTCCCAAAATGTTGGGATTACAGGTGAAGC 113741

QY 1256 CACTGACCCGGCTGATAGTGGTTTCATTTACTTACTTATTTCTTGACCACTCTGATCCATT 1315
|||||

Db 113740 CACTGACCCGGCTGATAGTGGTTTCATTTACTTACTTATTTCTTGACCACTCTGATCCATT 113681
|||||

QY 1316 TTGAAGTAAATGCTCCAAATTATATGCTGTTTGTAGAACACGGTAAGCATGTCATGTCG 1375
|||||

Db 113680 TTGAAGTAAATGCTCCAAATTATATGCTGTTTGTAGAACACGGTAAGCATGTCATGTCG 113621
|||||

QY 1376 TA---ATGCGCAGTGACATCAATAAAGAAAGTGCATTACTGAATGCTTTCAATGCTCTTA 1432
|||||

Db 113620 TAATTAATGCGCAGTGACATCAATAAAGAAAGTGCATTACTGAATGCTTTCAATGCTCTTA 113561
|||||

QY 1433 TAATGATGTAAGGTGGCATGTCATGGGCGCTATTAG-CCAGACATCACTCCAAAGAA 1491
|||||

Db 113560 TAATGATGTAAGGTGGCATGTCATGGGCGCTATTAGCCCGCAGACATCACTCCAAAGAA 113501
|||||

QY 1492 TTCCAAACAGATATAGCAAGTGCCTTTAGGGCCAGATCCCTTCCCTCAGCGCTGTTTA 1551
|||

Db 113500 TTCTNN 113441
|||

QY 1552 CCAGGGAATAGGATGCTCTGGGACAAGTTCCCTTAAGTGAAGTGTTCAT 1602
|||

Db 113440 NNN 113390
|||

RESULT 5

AL590622/c

LOCUS

DEFINITION

Human DNA sequence from clone RP11-320F15 on chromosome 10.
Contains the gene for ribonuclease P 30kD subunit, the gene for a
nuclear protein similar to CARP, ESTs, STSS and GSSs, complete
sequence.

ACCESSION

AL590622

VERSION

AL590622.7 GI:14270159

KEYWORDS

HTG; CARP; ribonuclease P.

SOURCE

Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 5011)

REFERENCE

Tracey,A.
Submitted (19-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk

AUTHORS

On May 31, 2001 this sequence version replaced gi:14161205.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.

JOURNAL

This sequence was finished as follows unless otherwise noted: all
regions were either double-stranded or sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality >=
30). An attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one M13 subclone; and the
assembly was confirmed by restriction digest. The following
abbreviations are used to associate primary accession numbers given
in the feature table with their source databases: Em: EMBL; Sw:
SWISSPROT; Tr: TREMBL; Wp: WORMPEP; Information on the WORMPEP
database can be found at
<http://www.sanger.ac.uk/Projects/C.elegans/wormpep> This sequence
was generated from part of bacterial clone contigs of human
chromosome 10, constructed by the Sanger Centre Chromosome 10
Mapping Group. Further information can be found at
<http://www.sanger.ac.uk/HGP/Chr10>

COMMENT

RP11-320F15 is from the library RPC1-11.2 constructed by the group
of Pieter de Jong. For further details see
<http://www.chori.org/bacpac/home.htm>

VECTOR: pBACE3.6

IMPORTANT: This sequence is not the entire insert of clone
RP11-320F15. It may be shorter because we sequence overlapping
sections only once, except for a short overlap.
The true left end of clone RP11-236B18 is at 50012 in this
sequence. The true right end of clone RP11-103A2 is at 100 in this
sequence.

Location/Qualifiers

1..50111

/organism="Homo sapiens"

/mol_type="genomic DNA"

/db_xref="taxon:9606"

/chromosome="10"

/clone="RP11-320F15"

/clone_lib="RPC1-11.2"

/complement(1..98)

/note="match: GSS: Em:AQ670367"

/complement(1..97)

/note="match: GSS: Em:AQ544636"

/complement(1..77)

/note="match: STS: Em:G56439"

/match: GSS: Em:AQ311708"

120..34469

/gene="BA320F15.1"

/product="BA320F15.1.1 (ribonuclease P (30kD) (RPP30))"

/note="match: CDNAS: Em:U77665 Em:AK004137 Em:U95123"

/match: ESTs: Em:BB572689 Em:BE378859 Em:BE137893"

Em:AA920575 Em:BG106820 Em:BF681617 Em:BF248001"

Em:BF031745 Em:AA854455 Em:A1359795"

/evidence=not experimental

join(391..472,3273)..3328,3962..4018,4428..4502,7467..7538,
14216..14305,23179..23295,23858..23887,24284..24321,
24717..24796,28974..29381)

/gene="BA320F15.1"

/product="BA320F15.1.1 (ribonuclease P (30kD) (RPP30))"

/note="match: CDNAS: Em:U77665 Em:AK004137 Em:U95123"

/match: ESTs: Em:BB572689 Em:BE378859 Em:BE137893"

Em:AA920575 Em:BG106820 Em:BF681617 Em:BF248001"

Em:BF031745 Em:AA854455 Em:A1359795"

/evidence=not experimental

join(391..472,3273)..3328,3962..4018,4428..4502,7467..7538,
14216..14305,23179..23295,23858..23887,24284..24321,
24717..24796,28974..29381)

/gene="BA320F15.1"

/note="match: proteins: Tr:P78346 Tr:O88796"

/codon_start=1

/evidence=not experimental

/product="BA320F15.1.1 (ribonuclease P (30kD) (RPP30))"

/protein_id="CAC70100.1"

/db_xref="GI:15717970"

/translation="MAVFADLDLRAGSLKALRGVETAHGLGYSVVAINHIVDFKEK
KQIEKPVAVSELTLPVVGKSPKILTLTLTIIVSDPSHCNVLRTSSRRLYDV
VAVFKTEKLPHIACTHLDVLVCITVEKLPFFYKRPPIVVAIRGLAFELVYSPAI
KDSMRRVTISSALNMQICGKNVITISSAARPLEIRGPDYDANLGLLGFUSEDAK
AAVSTNCRAALLHGHTRTAFGIISTVKPRPSEGDCLPASKAKCEG"

689..800

/note="MIR repeat: matches 30..132 of consensus"

1443..1488

/note="L2 repeat: matches 2459..2502 of consensus"

1761..2155

/gene="BA320F15.1"

/note="match: GSS: Em:AQ805524"

1907..2434

/gene="BA320F15.1"

/note="match: GSS: Em:AQ775672"

1926..2208

/gene="BA320F15.1"

/note="match: GSS: Em:B92129"

2349..2482

/note="L2 repeat: matches 2557..2696 of consensus"

2939..3167

/note="MIR repeat: matches 13..262 of consensus"

/complement(4176..4649)

/note="match: GSS: Em:AQ626500"

join(<4648..4764,7467..7538,14216..14305,23179..23887,
24284..24321,24717..24796,28974..29381)

/gene="BA320F15.1"

/product="BA320F15.1.2 (putative isoform 2)"

/note="match: ESTs: Em:AW939965"

```
/evidence=not_experimental
repeat_region 4733..4858
/note="L2 repeat: matches 2601..2732 of consensus"
repeat_region 4928..5102
/note="L2 repeat: matches 2299..2478 of consensus"
repeat_region 5257..5439
/note="L2 repeat: matches 1808..1971 of consensus"
repeat_region 5440..5721
/note="AluSx repeat: matches 7..293 of consensus"
repeat_region 5722..6237
/note="L2 repeat: matches 1200..1808 of consensus"
repeat_region 6884..7196
/note="AluJb repeat: matches 1..312 of consensus"
repeat_region 8243..8404
/note="2 copies 81 mer 81% conserved"
repeat_region 8445..8541
/note="HV1 repeat: matches 16..112 of consensus"
repeat_region 8542..8583
/note="HV1 repeat: matches 1..42 of consensus"
repeat_region 8950..9130
/note="AluSx repeat: matches 1..289 of consensus"
repeat_region 9601..9741
/note="MIR repeat: matches 76..212 of consensus"
repeat_region 10798..11076
/note="AluY repeat: matches 1..291 of consensus"
repeat_region 11086..11219
/note="FLAM_C repeat: matches 1..129 of consensus"
repeat_region 11319..11491
/note="MER5A repeat: matches 3..189 of consensus"
repeat_region 11928..11984
/note="L2 repeat: matches 1743..1801 of consensus"
repeat_region 13755..13817
/note="L2 repeat: matches 2684..2749 of consensus"
repeat_region 15390..15671
/note="AluJo repeat: matches 1..282 of consensus"
repeat_region 15791..16088
/note="AluSg repeat: matches 1..296 of consensus"
repeat_region 16979..17036
/note="L2 repeat: matches 2423..2481 of consensus"
misc_feature 17236..17755
/genes="BA320F15.1"
/note="match: GSS: Em:AQ551877"
repeat_region 17709..17972
/note="AluSx repeat: matches 45..304 of consensus"
repeat_region 19198..19363
/note="LIME repeat: matches 5272..5435 of consensus"
repeat_region 19489..19928
/note="MLT1C repeat: matches 5..478 of consensus"
repeat_region 21838..22258
/note="L2 repeat: matches 2109..2669 of consensus"
misc_feature 22440..22946
/genes="BA320F15.1"
/note="match: GSS: Em:AQ876596"
join(23179..23295,23858..23887,24284..24321,24717..24796,
28974..29072,30647..30748,31596..31680,33706..34469)
/genes="BA320F15.1"
/product="BA320F15.1.3 (putative isoform 3)"
/note="match: ESTs: Em:BE567341 Em:A1292002 Em:BF684192
Em:BF435407 Em:A1740881 Em:AA651912 Em:BF438915"
/evidence=not_experimental
misc_feature 24121..24518
/genes="BA320F15.1"
/note="match: GSS: Em:AQ026750"
complement(join(24284..24321,24724..24796,28974..29240))
/note="match: STS: Em:G24365"
misc_feature 24549..25020
/genes="BA320F15.1"
/note="match: GSS: Em:AQ223714"
misc_feature 24561..24956
/genes="BA320F15.1"
/note="match: GSS: Em:B53566"
repeat_region 25321..25458
/note="FLAM_C repeat: matches 1..133 of consensus"
```

```
misc_feature complement(25417..25863)
/note="match: GSS: Em:AQ712354"
repeat_region 26204..26603
/note="LIMA2 repeat: matches 5888..6302 of consensus"
repeat_region 28442..28516
/note="L2 repeat: matches 2636..2710 of consensus"
polyA_signal 29149..29154
/genes="BA320F15.1"
polyA_site 29166
/genes="BA320F15.1"

Query Match 32.1%; Score 664.8; DB 9; Length 50111;
Best Local Similarity 99.6%; Pred. No. 9.2e-141;
Matches 677; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY 1396 AAAGAAAAGTGAATCTACTGAATGCTTTCAATGTCTTATATATGATGTAAGTGGCATGTC 1455
DB 50111 AAAGAAAAGTGAATCTACTGAATGCTTTCAATGTCTTATATATGATGTAAGTGGCATGTC 50052

QY 1456 ATGGGGCCCTATTAG-CCCAGACATCACCTCCAAAGAAATTCCAAACAGATATAGACAAGTG 1514
DB 50051 ATGGGGCCCTATTAGCCCCCAGACATCACCTCCAAAGAAATTCCAAACAGATATAGACAAGTG 49992

QY 1515 CTTTATAGGGCCCATGATCCCTTCCCTCAGGCTGTTTACCAGGGAATAGGATGTCCTGGG 1574
DB 49991 CTTTATAGGGCCCATGATCCCTTCCCTCAGGCTGTTTACCAGGGAATAGGATGTCCTGGG 49932

QY 1575 ACAAGTTTCCCCTAAGTGAAGTGTGTAAGTCTGCTTATACAAAGATATTAATCTGGGG 1634
DB 49931 ACAAGTTTCCCCTAAGTGAAGTGTGTAAGTCTGCTTATACAAAGATATTAATCTGGGG 49872

QY 1635 TGTGATATCTAGGGCATCTACATTTTCTTGATAGGTAGTCAATATGAAGCTGCAAGAA 1694
DB 49871 TGTGATATCTAGGGCATCTACATTTTCTTGATAGGTAGTCAATATGAAGCTGCAAGAA 49812

QY 1695 AAAAGGGCAGTGAATGCTGCTGCAATGTCAACAGACAGCTGTCCCTGACTCTTGACAAAT 1754
DB 49811 AAAAGGGCAGTGAATGCTGCTGCAATGTCAACAGACAGCTGTCCCTGACTCTTGACAAAT 49752

QY 1755 AGGATGACTTGATCTGCTGAGCGATGTGATACCAACAAAGAAATGGCCCTCTCAATTT 1814
DB 49751 AGGATGACTTGATCTGCTGAGCGATGTGATACCAACAAAGAAATGGCCCTCTCAATTT 49692

QY 1815 CTTCTGATTCACATATTCAGCAGGGTGTAGCTTGCTCCCTCCCTCTTCAAGTTCCTCA 1874
DB 49691 CTTCTGATTCACATATTCAGCAGGGTGTAGCTTGCTCCCTCCCTCTTCAAGTTCCTCA 49632

QY 1875 GACACTGAGTCTGGAATGAAATTCACCTGCTCTGAGTTGGCTCCTAATGGGGGCGGA 1934
DB 49631 GACACTGAGTCTGGAATGAAATTCACCTGCTCTGAGTTGGCTCCTAATGGGGGCGGA 49572

QY 1935 GTGTTACTTCGGTTCCTCCAGTTGGAAGATTAATCTACCCGGCCCAAGCTATATAAGCTGA 1994
DB 49571 GTGTTACTTCGGTTCCTCCAGTTGGAAGATTAATCTACCCGGCCCAAGCTATATAAGCTGA 49512

QY 1995 CCGGTGTGAGGGGGCCACAGCGCCCACTCCAGGATTCCTTCCACGACAGAAACAT 2054
DB 49511 CCGGTGTGAGGGGGCCACAGCGCCCACTCCAGGATTCCTTCCACGACAGAAACAT 49452

QY 2055 ACAAGACTCTTTTCAGCCAAAC 2074
DB 49451 ACAAGACTCTTTTCAGCCAAAC 49432

RESULT 6
AX468603
LOCUS AX468603 2358 bp DNA linear PAT 16-JUL-2002
DEFINITION Sequence 1 from Patent WO0246220.
ACCESSION AX468603
VERSION AX468603.1 GI:21901402
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
```


Smit, A.P.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIBR
 Web site: <http://www-seq.wi.mit.edu>
 Contact: sequence_submissions@genome.wi.mit.edu

 Project Information
 Center project name: L25236
 Center clone name: 211_P_24

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 2 contigs. Gaps between the contigs
 * are represented as runs of N. The order of the pieces
 * is believed to be correct as given, however the sizes
 * of the gaps between them are based on estimates that have
 * provided by the submitter.
 * This sequence will be replaced
 * by the finished sequence as soon as it is available and
 * the accession number will be preserved.
 * 1 61126: contig of 61126 bp in length
 * 61127 61226: gap of 100 bp
 * 61227 142902: contig of 81676 bp in length.

FEATURES

Location/Qualifiers
 1..142902
 /organism="Mus musculus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10090"
 /clone="RP24-211P24"
 /clone.lib="RPCI-24 Male Mouse BAC"
 BASE COUNT 41376 a 30899 c 29987 g 40473 t 167 others

ORIGIN

Query Match 17.7%; Score 367.8; DB.2; Length 142902;
 Best Local Similarity 76.4%; Pred. No. 3.2e-73;
 Matches 597; Conservative 0; Mismatches 162; Indels 22; Gaps 11;
 QY 1289 TCATTCTTGACCACTCTGATCCATTTCGAAGTAAATGCTCCCAATATTATGCTGTT 1348
 Db 90844 TCCTCTGTGCATCACTTCGGCCGGTCTTGGGGT--AGATCCCTCTGATTAGCCTTCAGAT 90901
 QY 1349 TTAGAACACGGTAAGCATGTCATGTGCTA---ATGGCCAGTGCATCATATAAAGAAAGT 1405
 Db 90902 TTAGAACACGGTACGCTGTGGTGCACCTAATTATGCGCATGACACCATAGAGTCAAGT 90961
 QY 1406 GCATTACTGAATGCTTCAATGCTTTATTAATGATGTAAGGTGGCATGTACGGGCCCTA 1465
 Db 90962 GCATTACTGAATGCTTCAATGCTTTCTCTAATGCTGTGATGATGCGCATGTACAGGGCCAT 91021
 QY 1466 TTTAGC--CCAGACATCACTCAAGAAATTCACACAGATATACAGAGTGCCTTTAGGCG 1524
 Db 91022 TTTAGCTGCAGACATCACTCCAGAGAAATTCACACAGATAGACAGAGTGGCCACCCAGAC 91081
 QY 1525 CCAGATCCCTTCCCTCCAGCTGTGTACCCAGGGAATAGATGTCC---TGGGACCAAGTT 1581
 Db 91082 CCATCTCTTCCCTCCGCGTGATATATCCCAAGATAGATGTCCCAAGCAACACTTC 91141
 QY 1582 TCCCTTAAGTGAAGTGTGATAAGTCTGCTTATCAGAAAGATATTACTGGGGGTGTGATA 1641
 Db 91142 CCAGCCAACTGGAGTGTGATAAGTCCAGTTATCAGAAAGATATGCTGTGAAGTGTGATG 91201
 QY 1642 TGTAGGCGATCTACATTTTCTTGATA-GGTAGTCATATGAAGCTGACAAAGAA--AAAA 1698
 Db 91202 CACAGTGC--TTGCATTTTCTTGATACGTGTAGTCATATGAGAGCTGACAAAGAGAAAA 91259
 QY 1699 AGGGCAGTGTGTTGGTCAATGTCAACAGACAGCTGTCCCTCTGAC--TCTTGACAAATAGG 1757
 Db 91260 AGAGCAGCGATGTGGTGCATATTATACAGCAGCTGTCCCTCTGGCTTCCGATACGTGG 91319
 QY 1758 ATGACTTGCAATCTGAGCGATGTGATACCAACCAAGGAATGGCCCTCTCACATTTCTT 1817
 Db 91320 ATGACTGCAATCTGAGCGGTGTGCTACTGTCACAAAGGAATGACCCCTCTCACATTTCTT 91379

QY 1818 CTTGATTTCACATATTTCAGCAGGGTTAGTGTCTCTCCCTCCCTCTTCCAGTTCCTCCAGAC 1877
 Db 91380 CTTGATTTCACATACGCCGGG-----CCAGCTTGTCATCTCCCTCTTGGCTTCCAGAC 91434
 QY 1878 ACTGAGTCTGGAATGAAAAATTCACCTGCTCTGAGTTGGCTCTTAATGGGGCGGAGTG 1937
 Db 91435 ACTAAGTCTGGAATGAAAAATTCACCTGCTCTGAAATGGCCACTGCTGGTGGGCGGAGG 91494
 QY 1938 TTACTTCGGTTCCAGGTTGGAAGATTATCTCACCGGGCCCGAGCTATATAGTCAACG 1997
 Db 91495 TGACTTGGCTTCCAGGCTGGAAGATTATCTCACCGAGCCTAGCTATATAA--CGGGCTG 91553
 QY 1998 GTGTGAGGGGCCAGCAGCGGCCAACTCCAGGATTCCTTC--CACCACAGAAAAACATAC 2056
 Db 91554 GTGTGAGGGGCTCCACAGGGCCAGTTCAGGGGTTTCATCCACAGAGAGAAAAACATAG 91613
 QY 2057 A 2057
 Db 91614 A 91614

RESULT 8

AC105469 229640 bp DNA linear HTG 15-NOV-2002
 Rattus norvegicus clone CH230-140118, WORKING DRAFT SEQUENCE, 9
 unordered pieces.
 AC105469
 AC105469.4 GI:25007338
 HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
 Rattus norvegicus (Norway rat)
 SOURCE
 Rattus norvegicus
 ORGANISM
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 229640)
 Muzny, D., Maric, M., Metzker, M., Lee, S., Abramson, S., Adams, C., Alder, J.,
 Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
 Anyalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H.,
 Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
 Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
 Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
 Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Center, A.,
 Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J.,
 Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,
 Davila, M.L., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D.,
 Delgado, O., Denson, S., Deramo, C., Ding, Y., Dinh, H., Divya, K.,
 Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,
 Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G.,
 Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P.,
 Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M.,
 Gebregorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W.,
 Gunaratne, P., Haaland, W., Hamill, C., Hamilton, C., Hamilton, K.,
 Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J.,
 Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hoques, M.,
 Hollins, B., Howells, S., Hulyk, S., Hume, J., Idlebird, D., Jackson, A.,
 Jackson, L., Jacob, S., Kelly, S., Khan, Z., King, L., Kovar, C.,
 Karpathy, S., Kelly, S., Kelly, S., Johnson, B., Johnson, R., Jolivet, A.,
 Karpatis, S., Katt, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Kowis, C., Kratt, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J.,
 Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J.,
 Lorusshewa, L., Louleghed, H., Lozano, R.J., Lu, X., Ma, J.,
 Maheshwari, M., Mahindartne, M., Mahmoud, M., Malloy, K., Mangum, A.,
 Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E.,
 Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E.,
 Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S.,
 Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Naif, L.,
 Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S.,
 Nwakoelameh, O., Okwuonu, G., Olarnpunsagoon, A., Pal, S., Parks, K.,
 Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C.,
 Plopper, F., Polindexter, A., Popovic, D., Primus, E., Pu, L.,
 Puazo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R.,
 Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F.,
 Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S.J.,
 Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H.,

Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajls, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villaaana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausern, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission
Unpublished
2 (bases 1 to 229640)
Worley, K.C.

Direct Submission
Submitted (09-JAN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 229640)

Rat Genome Sequencing Consortium.
Submitted (15-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 15, 2002 this sequence version replaced gi:23101653.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GNAG
Center clone name: CH230-140118
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 215126 bases at least Q40
Consensus quality: 217436 bases at least Q30
Consensus quality: 218866 bases at least Q20
Estimated insert size: 214300; sum-of-contigs estimation
Quality coverage: 10x in Q20 bases; sum-of-contigs estimation

* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 215467: contig of 215467 bp in length
* 215468 215567: gap of unknown length
* 215568 218782: contig of 3215 bp in length
* 218783 218882: gap of unknown length
* 218883 220359: contig of 1477 bp in length
* 220360 220459: gap of unknown length
* 220460 221614: contig of 1155 bp in length
* 221615 221714: gap of unknown length
* 221715 223060: contig of 1346 bp in length

223061 223160: gap of unknown length
* 223161 224771: contig of 1611 bp in length
* 224772 224871: gap of unknown length
* 224872 225886: contig of 1015 bp in length
* 225887 225986: gap of unknown length
* 225987 228181: contig of 2194 bp in length
* 228181 228280: gap of unknown length
* 228281 229640: contig of 1360 bp in length.

FEATURES
Location/Qualifiers
source
1..229640
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-140118"
1..1268
/note="wgs_end_extension
clone_end:77"
4765..7050
/note="wgs_end_extension
clone_end:77"
7051..8156
/note="clone_boundary
clone_end:77"
8157..8156
end_sequence:RWAXF57TJB"
complement(7252..8118)
/note="clone_boundary
clone_end:77"
8119..215467
end_sequence:RWAXF57TJC"
/note="wgs_end_extension
clone_end:77"
215468..216791
215568..216791
/note="wgs_end_extension
clone_end:77"
216792..9702 others
BASE COUNT 61330 a 47195 c 48097 g 63316 t 9702 others
ORIGIN

Query Match 17.6%; Score 365; DB 2; Length 229640;
Best Local Similarity 77.3%; Pred. No. 1.4e-72;
Matches 601; Conservative 0; Mismatches 150; Indels 26; Gaps 12;

QY 1289 TCTATTTCTTGACCACTCTGATCCATTTTGAAGTAAATAATGCTCCAAATATTATATGCTGTT 1348
Db 141439 TGTGTCCTCGACCACTCTGACCAATTTGAGTCAATAT--TCCGATTAGCCTTCTGTT 141496

QY 1349 TTAGAACCGGTAAGCATGTCATGCTGAATGGCCAGTGACATCATATAAAGAAAAGTGCA 1408
Db 141497 TTAGAGCAC---ATGCCATGCACATAATATGGCCAGTGACACCATATAAAGTAAAGTGCA 141552

QY 1409 TTACTGAATGCTTTCAATGCTCTTATAATGATGCTAAGGTGGCATGTCATGGGGCCTATT 1468
Db 141553 TTACTGAATGCTTTCAATGCTCTTATAATGATGCTAAGGTGGCATGTCATGGGGCCTATT 141612

QY 1469 AGCCCG-AGACATCACTCCAAAGAAATTCACAAAGATATAGACAAGTGCCTTTAGGGCCCA 1527
Db 141613 AGCCCTGGACATCACTCCAGAGAAATTCACAAAGATATAGACAAGTGTCCACAGACCCA 141672

QY 1528 GATCCCTTCCCTCAGGCTGTTTA-CCAGGGAATAGGATGCTCTGGGACAAGTTTCCC- 1585
Db 141673 -----CTTCCCCCGGGCTGTTTATTCAGGAATAGGATGTCCCAAGAACACACTTCCA 141727

QY 1586 --CTAAGTGAAGTGTGATAGTCTGCTTATCAGAAAGATATCTAGGGGGTGTGATG 1643
Db 141728 GGCCCACTGGAGTGTGATAGTCTGATAGCCCAAGTATCAGAAAGATATTTGCCGTGAGTGTGATGCA 141787

QY 1644 TAGGGCATCTACATTTTCTTGATAGTAGTAGTATATGAAGAGTGCACAAAGAA--AAAAAGG 1701
Db 141788 CAATGCTTGCAC-TTTCCTGATAGGTAGTCTATACGAAGCTGACAGAGAGGAGGAAAGG 141846

QY 1702 GCAGTGATGTGGTCAATGTCAACAGACAGCTGTCCCTCTACTCTTTGACAAATAGGATGA 1761
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||

```
Db 141847 GCAGCGATGTGGTCAATATGAACAGCAGCGTGTCTCTGGCTTCCCGATAAGTAGGATGA 141906
QY 1762 CTTGCAATGTGTGACGATGTGATCATCACCAAGGAATGGCCCTCTCACATTTCTTCCTG 1821
Db 141907 CTCGCATTTGTGGCGGTGTGGTCACTGCCAATGGAATGCCCTCTCACATTTCTTCCTG 141966
QY 1822 ATTTCACATATTACAGCAGGGTTAGCTTGCTCTCCCTCCCTCTTCAGCTTCCACGACACTG 1881
Db 141967 ATTTCGCACA--CACACGGCCAGCTGTGTAT---CTCCCTCTTGGCTTCCCGACACACTA 142021
QY 1882 AGTCTGGAATGAAATTCACCTGCCTCTGAGTTGGTCTCTTAATGGGGGGGAGTGTTCAC 1941
Db 142022 AGTCTGGAATGAAATTCACCTGCCTCTGAATTGGCCACAGCGAGGCGGGTGTGCAC 142081
QY 1942 TTCCGTTTCCAGGTTGGAATTTATCTACCCGGCCCCAGCTATATAAGCTGACCGGTGT 2001
Db 142082 TTGGCTTCCAGGCTGGAATTTATCTACCCAGTCTAGTATATAAG-AGGCTGGGT 142140
QY 2002 GGAGGGGCCAGCAGGCGCAACTCCAGGANTCTTTC-CAGCACAAACATACA 2057
Db 142141 GGAGGGGCTCCACAGGCGCAGTCCAGGGTTCAGCCACAAGAGGGAACATAGA 142197

RESULT 9
AC097115
LOCUS 238344 bp DNA linear HTG 14-NOV-2002
DEFINITION Rattus norvegicus clone CH230-26A2, *** SEQUENCING IN PROGRESS ***,
2 unordered pieces.
AC097115
VERSION AC097115.6 GI:24956605
KEYWORDS HTG; HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 238344)
Muzny,D.Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
Cardenas,V., Carter,K., Cavazos,I., Cesar,H., Center,A.,
Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
Davilla,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,
Hollins,B., Howells,S., Huliyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpach,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensnewa,L., Loulesged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
Morean,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
Paokelemeh,O., Okunou,G., Olarnpusagoon,A., Pal,S., Pfannkuch,C.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puzo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
```

```

Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvarcsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneddy,A., Sodargren,B., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wiczzyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.
Direct Submission
Unpublished
2 (bases 1 to 238344)
Worley,K.C.
Direct Submission
Submitted (11-OCT-2001) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 238344)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (14-NOV-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Nov 14, 2002 this sequence version replaced gi:22855482.
The sequence in this assembly is a combination of BAC based reads
and whole genome shotgun sequencing reads assembled using Atlas
(http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
in the feature table below represents a scaffold in the Atlas
assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the clone and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: http://www.hgsc.bcm.tmc.edu/
Contact: hgsc-help@bcm.tmc.edu
----- Project Information
Center project name: GGOJ
Center clone name: CH230-26A2
----- Summary Statistics
Assembly program: Phrap; version 0.990329
Consensus quality: 222569 bases at least Q40
Consensus quality: 226638 bases at least Q30
Consensus quality: 229427 bases at least Q20
Estimated insert size: 233018; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: Estimated insert size may differ from sequence length
* (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 236546: contig of 236546 bp in length
* 236547 236546: gap of unknown length
* 236647 238344: contig of 1698 bp in length.
Location/Qualifiers
1. 238344
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
FEATURES
source
```

ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS	1 (bases 1 to 723)
TITLE	Maeda,T., Sepulveda,J., Chen,H.H. and Stewart,A.F.R. alphas-Adrenergic activation of the cardiac ankyrin repeat protein gene in cardiac myocytes
JOURNAL	Gene 297 (1-2), 1-9 (2002)
REFERENCE	2 (bases 1 to 723)
AUTHORS	Maeda,T., Sepulveda,J. and Stewart,A.F.R.
TITLE	Direct Submission
JOURNAL	Submitted (29-JAN-2002) Cardiovascular Institute, University of Pittsburgh, 200 Lothrop Street, Pittsburgh, PA 15213, USA
FEATURES	Location/Qualifiers
source	1..723
	/organism="Mus musculus"
	/mol_type="genomic DNA"
	/strain="FVB"
	/db_xref="taxon:10090"
gene	1..>723
	/gene="Carp"
misc_feature	1..720
	/gene="Carp"
	/note="contains promoter and 5' UTR"
mRNA	<721..>723
	/gene="Carp"
	/product="cardiac ankyrin repeat protein"
CDS	721..>723
	/gene="Carp"
	/codon_start=1
	/product="cardiac ankyrin repeat protein"
	/protein_id="AAI85342.1"
	/db_xref="GI:19110907"
	/translation="M"
BASE COUNT	187 a 185 c 180 g 171 t
ORIGIN	
Query Match	16.4%; Score 340.8; DB 10; Length 723;
Best Local Similarity	74.0%; Pred. No. 5.2e-67;
Matches 527; Conservative	0; Mismatches 172; Indels 13; Gaps 7;
QY	1352 GAACACGGTAAAGCATGTTCATGTGCTA---ATGGCCAGTGCATCATATAAAGAAAGTGC 1408
Db	1 GAACACGGTAAAGCATGTTCATGTGCTA---ATGGCCAGTGCATCATATAAAGAAAGTGC 60
QY	1409 TTACTGAATGCTTTCAATGTCTTATATGATGGTAAAGTGGCATGTTCATGGGCTATTTT 1468
Db	61 TTACTGAATGCTTTCAATGTCTTATATGATGGTAAAGTGGCATGTTCATGGGCTATTTT 120
QY	1469 AGC-CCAGACATCACTCCAAAGAATTCACAAACAGATATAGCAAGTGCCTTTTAGGGCCCA 1527
Db	121 AGCTGCAGACATCACTCCAGAGAATTCACAAACAGATAGACAGTGGCACCAGACCCA 180
QY	1528 GATCCCTTCCCTCAGGCTGTTTACCAGGGAATAGGATGTCCTGGGACAAAGTTTCCCTT 1587
Db	181 TCTCCTTCCCTCAGGCTGTTTACCAGGGAATAGGATGTCCTGGGACAAAGTTTCCCTCAG 240
QY	1588 AAGTCAAGTGTGTAAAGTGTCTTATCAGAAAGATATTACTGGGGTGTGATATGTAGG 1647
Db	241 AACTGGATGCGGTAAAGTCCAGTTATCAGAAAGATATGGCTGTAAAGTGTGATCGACAGT 300
QY	1648 GCATCTACATTTTCTTGATAGGTAGTTCATATGAAGCTGCACAAAGAAAGAAAGGCGAGT 1707
Db	301 GCTTCA-TTTTTCTTGATAGGTAGTTCATATGAAGCTGCACAAAGAAAGAAAGGCGAGT 359
QY	1708 ATGTGGTGCATATGTCAACAGACAGCTGTCCCTGAC-TCTTGAACAATAGGATGACTTGC 1766
Db	360 GATGTGTCAATATTAACAGGCAGCTGTCCCTGGCTTCCCGATACGTGGATGACTCGC 419
QY	1767 ATTGCTGAGCGATGTGATCACCACAAAGGAATGGCCCTCTCACATTTCTTCTGATTCA 1826
Db	420 ATTGCTGAGCGGTGTGGTCACTGCCAAGGAATGACCCCTCTCAATTTCTTCTGATTTC 479

[illegible]

JOURNAL MEDLINE PUBMED REFERENCE AUTHORS TITLE JOURNAL

nuclear protein from human endothelial cells
J. Biol. Chem. 270 (17), 10236-10245 (1995)
95247734
7730328
2 (bases 1 to 1901)
Chu, W.
Direct Submission
Submitted (05-JAN-1995) W. Chu, Hoffmann-La Roche, 340 Kingsland
Street, Dept. of Inflammation/Autoimmune Disease, Hoffmann-La
Roche, Nutley, NJ 07110, USA

FEATURES
source
1..1901
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="C-193"
/cell_type="endothelial"
/tissue_type="skin"
/clone_lib="HDMEC cDNA"
1..1901
94..98
/note="nuclear localization signal"
152..283
/note="ankyrin-like repeats"
250..1209
/note="cytokine-inducible expression"
/codon_start=1
/product="nuclear protein"
/protein_id="CAA58676.1"
/db_xref="GI:793841"
/translation="MMVLKVEELVTGKNGNGEAGFLPEDFDGEYEAATLEKQED
LKTLLAHPVTGLGQWKSQKAEELPKKLEORSKLENLELIIQLKKKVKRKT
KYPVVKPEPEPIITEPVDVPTLKALENKLPVVEKFLSKNPNPVCDEYKKTALHRA
CLEGHAIIVEKLMEGAQIEFRDMLSESTAIHWASRGNDVDKLLNKGAKISARDKL
LSTALHVAVRTGHYCAEHLIAEADLNADREGDTPPLHDVRLNRYKMRILLIYGA
DLINKACGKTPMDLVLHWQNGTKAIFDSLRENSYKTSRIATF"
BASE COUNT 592 a 378 c 460 g 471 t
ORIGIN

Query Match 11.0%; Score 228; DB 9; Length 1901;
Best Local Similarity 97.6%; Pred. No. 2.3e-41;
Matches 242; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1828 ATATTACAGAGGTTAGCTTGT-CCTCCCTCCCTCTTCAGCTTCCACACACTGAGTCT 1886
|||
Db 2 AAAAAACAGAGGTTAGCTTGTCCCTCCCTCTTCAGCTTCCACACACTGATTTCT 61
|||

QY 1887 GGAATGAAATTCACCTGCTCTGAGTTGGCTCCTAATGGGGCGGAGTGTACTTCGG 1946
|||
Db 62 GGAATGAAATTCACCTGCTCTGAGTTGGCTCCTAATGGGGCGGAGTGTACTTCGG 121
|||

QY 1947 TTCCCAGGTTGGAAGATTATCTACCCGCCCGCCAGCTATATAAGCTGACCGGTGGAGG 2006
|||
Db 122 TTCCCAGGTTGGAAGATTATCTACCCGCCCGCCAGCTATATAAGCTGACCGGTGGAGG 181
|||

QY 2007 GGGCCAGCAGGCGCCAACTCCAGGATTCTTCCACGACAGAAAAACATACAGACTCCTT 2066
|||
Db 182 GGGCCAGCAGGCGCCAACTCCAGGATTCTTCCACGACAGAAAAACATACAGACTCCTT 241
|||

QY 2067 CAGCCCAAC 2074
|||
Db 242 CAGCCCAAC 249
|||

RESULT 14
G28603
LOCUS G28603 1901 bp DNA linear STS 11-JUL-1996
DEFINITION human STS SHGC-35401, sequence tagged site.
ACCESSION G28603
VERSION G28603.1 GI:1408418
KEYWORDS STS; STS sequence; primer; sequence tagged site.

SOURCE ORGANISM

REFERENCE AUTHORS JOURNAL COMMENT

Homo sapiens (human)
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1901)
Myers, R.M.
Unpublished (1996)

Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 4157259687
Fax: 4157259689
Email: myers@shgc.stanford.edu

Primer A: GGCATTTTGAAGGCATGG
Primer B: CCAGATGATGATCATGAAG
STS size: 222
PCR Profile:

Initial incubation: 94 degrees C for 90 seconds

Denaturation: 94 degrees C for 15 seconds
Annealing: 62 degrees C for 23 seconds
Polymerization: 72 degrees C for 30 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9600

Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
Tag Polymerase: 0.05 units/uL
Total Vol: 10 uL

Buffer:
MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 20 mM
pH: 8.3

Prepared with primer pairs provided by Sandoz, derived from X83703
-- Washington University/Merck EST sequence.

FEATURES
source

1..1901
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/map="10"

STS
1246..1467

primer_bind 1246..1263

primer_bind complement(1447..1467)

BASE COUNT 592 a 378 c 460 g 471 t

ORIGIN

Query Match 11.0%; Score 228; DB 11; Length 1901;
Best Local Similarity 97.6%; Pred. No. 2.3e-41;
Matches 242; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

QY 1828 ATATTACAGAGGTTAGCTTGT-CCTCCCTCCCTCTTCAGCTTCCACACACTGAGTCT 1886

Db 2 AAAAAACAGAGGTTAGCTTGTCCCTCCCTCTTCAGCTTCCACACACTGATTTCT 61

QY 1887 GGAATGAAATTCACCTGCTCTGAGTTGGCTCCTAATGGGGCGGAGTGTACTTCGG 1946

Db 62 GGAATGAAATTCACCTGCTCTGAGTTGGCTCCTAATGGGGCGGAGTGTACTTCGG 121

QY 1947 TTCCCAGGTTGGAAGATTATCTACCCGCCCGCCAGCTATATAAGCTGACCGGTGGAGG 2006

Db 122 TTCCCAGGTTGGAAGATTATCTACCCGCCCGCCAGCTATATAAGCTGACCGGTGGAGG 181

QY 2007 GGGCCAGCAGGCGCCAACTCCAGGATTCTTCCACGACAGAAAAACATACAGACTCCTT 2066

Db 182 GGGCCAGCAGGCGCCAACTCCAGGATTCTTCCACGACAGAAAAACATACAGACTCCTT 241

QY 2067 CAGCCAAC 2074
Db 242 CAGCCAAC 249

RESULT 15
AF131883
LOCUS 1940 bp mRNA linear MAM 30-NOV-1999
DEFINITION Oryctolagus cuniculus CARP mRNA, complete cds.
ACCESSION AF131883
VERSION AF131883.1 GI:6478316
KEYWORDS
SOURCE Oryctolagus cuniculus (rabbit)
ORGANISM Oryctolagus cuniculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
REFERENCE 1 (bases 1 to 1940)
AUTHORS Aihara, Y., Kurabayashi, M., Arai, M., Kedes, L. and Nagai, R.
TITLE Molecular cloning of rabbit CARP cDNA and its regulated expression
in adriamycin-cardiomyopathy
JOURNAL Biochim. Biophys. Acta 1447 (2-3), 318-324 (1999)
MEDLINE 20011295
PUBMED 10542334
REFERENCE 2 (bases 1 to 1940)
AUTHORS Aihara, Y.
TITLE Direct Submission
JOURNAL Submitted (26-FEB-1999) Second Department of Internal Medicine,
Gunma University School of Medicine, 3-39-15 Showa-machi, Maebashi
371-8511, Japan

FEATURES
source 1..1940
/organism="Oryctolagus cuniculus"
/mol_type="mRNA"
/db_xref="taxon:9986"
219..1178
/codon_start=1
/product="CARP"
/protein_id="AAFI3817.1"
/db_xref="GI:6478317"
/translation="MMVLVEELVTGKSSNGETGEFLPDDFRDGOYEAAVTSEKQED
LKTLPAAHVSLEAEQAELEKKLEQSKLENLELEIIOLKKRYKRT
KVPVAKPEPEIITEPVDVPRFLKALENKLAVEKFLSDONNPDVCDEVKRTALHRA
CLEGLAIIVEKLEMAQAQIEPRDMLBSTAIHWACGGNLEVLKLLINKGAKISARDKL
LSTALHVAVRTGHECAEHLIACEADLNADREGDTPFLHDAVRLNRYKMRILLIMTGA
DITIKNSAGKTPMDLVNWNQNGTKAIFDSLKENSXYKTSRIATF"

BASE COUNT 623 a 391 c 479 g 447 t
ORIGIN

Query Match 7.4%; Score 154.4; DB 4; Length 1940;
Best Local Similarity 84.9%; Pred. NO. 1.3e-24;
Matches 185; Conservative 0; Mismatches 31; Indels 2; Gaps 1;

QY 1859 CCTCTTCAGCTTCCACAGACACTGAGTCTGGAATGAAATTCACCTGCCTCTGAGTTGGCT 1918
Db 1 CCTCTTGGGCTTCCACAGACACTGAGTCTGGAATGAAATTCACCTGCCTCTGAGTTGGCT 60

QY 1919 CCT--AATGGGGGGGAGTGTACTTCGGTTCACAGTTGGAAGATTATCTACCCGGC 1976
Db 61 ACTGGCAGGGGGGAGGGCTGTACTTGGTTCCAGGTTGGAAGATTATCTACCCAGC 120

QY 1977 CCCAGTATATAGCTGACCGGTGTGGAGGGGCCAGCGGCAACTCCAGGATTCTT 2036
Db 121 CCAAACTATATAGCCGAGCAGTGTGGAGGGGACAGCAGCGGCCATCTCCAGGGATTCT 180

QY 2037 TCCACGACAGAAAAACATACAGACTCTTCAGCCAAC 2074
Db 181 TCCACAGGAGAAAAACACACAGCGACTCCAGCCAAC 218

Search completed: November 18, 2003, 04:22:10
Job time : 7557.42 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 23:28:30 ; Search time 543.77 Seconds
(without alignments)
10295.963 Million cell updates/sec

Title: US-10-005-337A-2

Perfect score: 2074

Sequence: 1 ctgcagaaagtacttaagt.....acaagactcttcagccaac 2074

Scoring table: IDENTITY NUC

Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2071.6	99.9	2074	24	ABV73021 Human CARP protein
2	367.8	17.7	2358	24	ABV73020 Mouse CARP protein
3	322.6	15.6	2247	21	AA10406 Murine cardiac ank
4	229.6	11.1	1988	22	AA193508 Human polynucleoti
5	228	11.0	1901	22	AAH02910 Human shear stress
6	228	11.0	1901	22	AAH02910 Human CAAS8676 pro
7	144.4	7.0	1889	24	AA594903 Human DNA sequence
8	144.4	7.0	1889	25	ABX77648 Differentially exp

ALIGNMENTS

RESULT 1
ABV73021

ID ABV73021 standard; DNA; 2074 BP.

XX AC ABV73021;

XX DT 08-JAN-2003 (first entry)

XX DE Human CARP protein coding sequence upstream DNA fragment.

XX KW Cardiac ankyrin repeat protein; CARP; cardiac; immunosuppressive;
XX KW antiinflammatory; Gene therapy; antisense gene therapy; human; ds.

XX OS Homo sapiens.

XX PN WO200246220-A2.

XX PD 13-JUN-2002.

XX PF 05-DEC-2001; 2001WO-EP15412.

XX PR 07-DEC-2000; 2000US-251582P.

XX PA (AVET) AVENTIS PHARMA SA.
(REGC) UNIV CALIFORNIA.

XX PA (BENO/) BENOIT P.

XX PI Schwartz B, Branellec D, Chien K;

XX DR Human DNA sequence
WPI; 2002-740642/80.

Human neuroblastom
Human kinesin-like
Human immune/haema
Genomic sequence #
Human immune/haema
Human immune/haema
Human immune/haema
Human immune/haema
Human nervous syst
Human nervous syst
Human immune/haema
Human immune/haema
Human musculoskele
cDNA encoding nove
Human cDNA differe
Human TBC-1 partia
Genomic sequence #
Human immune/haema
Human digestive sy
Human nervous syst
Human transporter
Human immune/haema
Human immune/haema
Human immune/haema
Human CGMP-inhibit
Human immune/haema
Human hypoxanthine
Human immune/haema
Human immune/haema
Human immune/haema
Human autoimmune d
Human secreted pro
Human musculoskele
cDNA encoding nove
Human immune/haema
Human immune/haema
Human cDNA differe
Human D-amino acid

PT	New promoter sequence derived from a portion upstream of the coding
PT	sequence of a gene for Cardiac Ankyrin Repeat Protein, for controlling
PT	the level and specificity of expression of a transgene in cardiac
PT	muscle cells
XX	
PS	Claim 5; Fig 2; 48pp; English.
XX	
CC	The invention relates to a polynucleotide (I) comprising a fragment of a
CC	sequence upstream of the coding part of the gene for the Cardiac Ankyrin
CC	Repeat Protein (CARP). (I) is capable of inducing a specific expression
CC	in vivo of a gene operably linked to (I), in cardiac cells. (I) or a
CC	vector (Iib) comprising (I) is useful for the manufacture of a medicament
CC	intended for the treatment of cardiac insufficiency, cardiac hypertrophy
CC	and hypoxia, and for preventing rejection during cardiac transplant. An
CC	expression cassette under the control of (I) is useful for encoding a
CC	protein or RNA which is capable of activating the growth of cardiac
CC	cells, reducing or suppressing an immune response, inducing angiogenesis,
CC	correcting muscle contractility, cardiac hypertrophy, cardiac
CC	insufficiency and myocarditis. (Iib) is useful for expressing a gene of
CC	therapeutic interest in vivo, by isolating (Iib) and introducing (Iib) in
CC	the cardiac tissue, under conditions so that the gene of interest is
CC	expressed. (I), the vectors and the compositions are useful in clinical,
CC	experimental, therapeutic and diagnostic fields, and in the treatment and
CC	prevention of cardiac pathologies. (I) is also useful for generating
CC	transgenic animals which constitute models for studying certain cardiac
CC	pathologies. The transgenic animals are also useful for screening
CC	molecules for their activity on the regulatory sequences of the gene
CC	encoding the CARP protein. The present sequence represents the DNA
CC	fragment upstream of the coding sequence of a human CARP protein.
XX	
SQ	Sequence 2074 BP; 612 A; 469 C; 416 G; 572 T; 5 other;
	Query Match 99.9%; Score 2071.6; DB 24; Length 2074;
	Best Local Similarity 100.0%; Pred. No. 0;
	Matches 2074; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 CTCGAGCAAGTTACTTAATGTTTTTTCCTCAGCATCCTCTCTGTAATAATGAGAGCATTA 60
DB	1 CTCGAGCAAGTTACTTAATGTTTTTTCCTCAGCATCCTCTCTGTAATAATGAGAGCATTA 60
QY	61 GTCTTGTCTCAACTTCGAGGGCATGGACAGCTCTGGGATTTTCATATCCAGACCCTTAAA 120
DB	61 GTCTTGTCTCAACTTCGAGGGCATGGACAGCTCTGGGATTTTCATATCCAGACCCTTAAA 120
QY	121 CATCCACAGTCTTCCCCCAAAACACTTCTCTCTTAATACCTCCCTCAGTTTGGGTGAG 180
DB	121 CATCCACAGTCTTCCCCCAAAACACTTCTCTCTTAATACCTCCCTCAGTTTGGGTGAG 180
QY	181 GCTGGAAACAAAAGGCATACGAATGGTAGAAAAGTGTCATGCTACTCTTGACTTA 240
DB	181 GCTGGAAACAAAAGGCATACGAATGGTAGAAAAGTGTCATGCTACTCTTGACTTA 240
QY	241 GATGAAGAGACCAATGAAAATAGTAATGACTCTGTGTTTGTCTCAGCAGGACATATCTAAA 300
DB	241 GATGAAGAGACCAATGAAAATAGTAATGACTCTGTGTTTGTCTCAGCAGGACATATCTAAA 300
QY	301 ATAGGAGCTATACAAAGAAATAGTACATGGACTCTGTGCAAGAATGACACACAAATTTGT 360
DB	301 ATAGGAGCTATACAAAGAAATAGTACATGGACTCTGTGCAAGAATGACACACAAATTTGT 360
QY	361 GAAACATTCATATATTAATAATAATAATAATAAGAAAAGGAAAATAATAAAG 420
DB	361 GAAACATTCATATATTAATAATAATAATAATAAGAAAAGGAAAATAATAAAG 420
QY	421 AAAATAGTAGTAGTGTGTCTCACTCAAGAAAAGCCAGGAGATTTCTTTATTTACCC 480
DB	421 AAAATAGTAGTAGTGTGTCTCACTCAAGAAAAGCCAGGAGATTTCTTTATTTACCC 480
QY	481 CTTTTAAGATAGAAATATAGGAGACCGGAACATATGATACAGAGGTACTGGAGGTCC 540
DB	481 CTTTTAAGATAGAAATATAGGAGACCGGAACATATGATACAGAGGTACTGGAGGTCC 540
QY	541 CTCCTTGTCAATGTTTTTCTTGGGGTGGGAGTCGATGCTCTCAAGTTTCAGAAAC 600

541	CTCTTGTCAATGTTTTTCTTGGGGTGGGAGTCGATGCTCTCTCAAGTTTCAGAAAC	600	
601	ACCATCCACTGACTGAGCAATTCAGGGGCAAGAGGAATGCGAGCCACATTTGTTGATT	660	
601	ACCATCCACTGACTGAGCAATTCAGGGGCAAGAGGAATGCGAGCCACATTTGTTGATT	660	
661	GGGTGAGTTTGGGAGAAATAGACACAAAGGTCAAACTAACTTCTCTAAATTAACACTT	720	
661	GGGTGAGTTTGGGAGAAATAGACACAAAGGTCAAACTAACTTCTCTAAATTAACACTT	720	
721	CCCTCCATTCACAAATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	780	
721	CCCTCCATTCACAAATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	780	
781	AGTTTTCCTGAACTATATAAATACCCCGAGTATGTTTACATAATTTACACCTCAAGA	840	
781	AGTTTTCCTGAACTATATAAATACCCCGAGTATGTTTACATAATTTACACCTCAAGA	840	
841	TTAGAAACAGAAATAGACACCTTTTCAACCTTCCGGAAGCAAGAGTCATATCCCTCC	900	
841	TTAGAAACAGAAATAGACACCTTTTCAACCTTCCGGAAGCAAGAGTCATATCCCTCC	900	
901	AGCCAGTCTCTCAAAATCTTGATGATCAGATCAATCATCTGGTGCTTTKAAATTCAGAGT	960	
901	AGCCAGTCTCTCAAAATCTTGATGATCAGATCAATCATCTGGTGCTTTKAAATTCAGAGT	960	
961	ATTCTTACAGTTTACATAAATCACTCAGAAATTCCTCGAGTGGGGCAGGGATCTGTA	1020	
961	ATTCTTACAGTTTACATAAATCACTCAGAAATTCCTCGAGTGGGGCAGGGATCTGTA	1020	
1021	TTTCTGACAGCTCCACAGGTGATTCTTCCACACAGCATTTGAGAACTTCAGCTCAA	1080	
1021	TTTCTGACAGCTCCACAGGTGATTCTTCCACACAGCATTTGAGAACTTCAGCTCAA	1080	
1081	TGACCTTAATCAGAGTCTCTCCATTTGCTAATATCTGCTCTCATTTTBTCTATATATA	1140	
1081	TGACCTTAATCAGAGTCTCTCCATTTGCTAATATCTGCTCTCATTTTBTCTATATATA	1140	
1141	TAGTATTTGTGTAGAGATGGGATTTTGGCATTTGCCAGGCTAGTATTGAATCTCTAA	1200	
1141	TAGTATTTGTGTAGAGATGGGATTTTGGCATTTGCCAGGCTAGTATTGAATCTCTAA	1200	
1201	GCTAAGCAATCTTCTGCTCTCTGCTCCCAAAATGTTGGGATTTACAGGTGTAAGCCACTG	1260	
1201	GCTAAGCAATCTTCTGCTCTCTGCTCCCAAAATGTTGGGATTTACAGGTGTAAGCCACTG	1260	
1261	CACCCGGCTGATAGCTGGTTTCATTTTACTCTATTTTTCACCACTCTGATCCATTTTGA	1320	
1261	CACCCGGCTGATAGCTGGTTTCATTTTACTCTATTTTTCACCACTCTGATCCATTTTGA	1320	
1321	GTAAATAATCTCCAATTTATTTATGCTTTTGAACACCGTAAAGTGTCAATGCTTAATG	1380	
1321	GTAAATAATCTCCAATTTATTTATGCTTTTGAACACCGTAAAGTGTCAATGCTTAATG	1380	
1381	GCAGTGCATCATATAAAGAAAAGTGCAATCTGTAATGCTTTCAATGCTCTTATAATG	1440	
1381	GCAGTGCATCATATAAAGAAAAGTGCAATCTGTAATGCTTTCAATGCTCTTATAATG	1440	
1441	GTAAAGTGGCATGCTCATGGGGCTATTATAGCCAGACATCACTCCAAAAGAAATTCAAA	1500	
1441	GTAAAGTGGCATGCTCATGGGGCTATTATAGCCAGACATCACTCCAAAAGAAATTCAAA	1500	
1501	GATATAGCAAGTGGCTTTTAGGGCCAGATCCCTTCCCTCAGGCTGTTTACCAGGAA	1560	
1501	GATATAGCAAGTGGCTTTTAGGGCCAGATCCCTTCCCTCAGGCTGTTTACCAGGAA	1560	
1561	TAGGATGCTCTGGGACAAAGTTTCCCTTAAGTGAAGTTGTAAGTCTGCTTATCAGAA	1620	
1561	TAGGATGCTCTGGGACAAAGTTTCCCTTAAGTGAAGTTGTAAGTCTGCTTATCAGAA	1620	
1621	GATATTACTTGGGGTGTGATATGTAGGCACTCTACATTTTCTTGATAGGTAGTCATATGA	1680	

Db 1621 GATATTACTGGGGTGTGATATGATAGGGAATCTACATTTTCTTGATAGGTAGTCATATGA 1680
 Qy 1681 AAGCTGACAAAGAAAAAGGGCAGTGATGGTGCATATGTCGCAATGTCACAGACAGCTGTCCCT 1740
 Db 1681 AAGCTGACAAAGAAAAAGGGCAGTGATGGTGCATATGTCGCAATGTCACAGACAGCTGTCCCT 1740
 Qy 1741 GACTCTTGACAAATAGATGACTTGCATGCTGAGGATGTCATCACCACCAAGGAATG 1800
 Db 1741 GACTCTTGACAAATAGATGACTTGCATGCTGAGGATGTCATCACCACCAAGGAATG 1800
 Qy 1801 GCCTCTCACATTTCTTCTGATTCACATATTCAGCAGGTTAGTTCCTCCCTCC 1860
 Db 1801 GCCTCTCACATTTCTTCTGATTCACATATTCAGCAGGTTAGTTCCTCCCTCC 1860
 Qy 1861 TCTCAGCTTCCAGACACTGAGTCTGGAATGAAATTCACCTGCTCTGAGTGTGCTCC 1920
 Db 1861 TCTCAGCTTCCAGACACTGAGTCTGGAATGAAATTCACCTGCTCTGAGTGTGCTCC 1920
 Qy 1921 TAATGGGGGGGAGTGTACTTCGGTTCCTCAGGTTGGAAGATATCTCACCGGCCCA 1980
 Db 1921 TAATGGGGGGGAGTGTACTTCGGTTCCTCAGGTTGGAAGATATCTCACCGGCCCA 1980
 Qy 1981 GCTATATAAGCTGACCGGTGTGGAGGGCCAGCAGGCGCAACTCCAGGGATTCTCTTCCA 2040
 Db 1981 GCTATATAAGCTGACCGGTGTGGAGGGCCAGCAGGCGCAACTCCAGGGATTCTCTTCCA 2040
 Qy 2041 CGACAGAAAAACATAAAGACTCTTTCAGCCAA 2074
 Db 2041 CGACAGAAAAACATAAAGACTCTTTCAGCCAA 2074

RESULT 2

ABV73020
 ID ABV73020 standard; DNA; 2358 BP.
 AC ABV73020;
 XX
 DT 08-JAN-2003 (first entry)
 DE Mouse CARP protein coding sequence upstream DNA fragment.
 DE
 KW Cardiac ankyrin repeat protein; CARP; cardiac; immunosuppressive;
 KW antiinflammatory; gene therapy; antisense gene therapy; mouse; ds.
 XX
 OS Mus musculus.
 XX
 PN WO200246220-A2.
 XX
 PD 13-JUN-2002.
 XX
 PF 05-DEC-2001; 2001WO-EP15412.
 XX
 PR 07-DEC-2000; 2000US-251582P.
 XX
 PA (AVET) AVENTIS PHARMA SA.
 PA (REGC) UNIV CALIFORNIA.
 PA (BENO/) BENOIT P.
 XX
 PI Schwartz B, Branellec D, Chien K;
 XX
 DR WPI; 2002-740642/80.
 XX
 PT New promoter sequence derived from a portion upstream of the coding
 PT sequence of a gene for Cardiac Ankyrin Repeat Protein, for controlling
 PT the level and specificity of expression of a transgene in cardiac
 PT muscle cells -
 XX
 PS Claim 1; Fig 1; 48pp; English.
 XX
 CC The invention relates to a polynucleotide (I) comprising a fragment of a
 CC sequence upstream of the coding part of the gene for the Cardiac Ankyrin
 CC Repeat Protein (CARP). (I) is capable of inducing a specific expression
 CC in vivo of a gene operably linked to (I), in cardiac cells. (I) or a

CC vector (Iib) comprising (I) is useful for the manufacture of a medicament
 CC intended for the treatment of cardiac insufficiency, cardiac hypertrophy
 CC and hypoxia, and for preventing rejection during cardiac transplant. An
 CC expression cassette under the control of (I) is useful for encoding a
 CC protein or RNA which is capable of activating the growth of cardiac
 CC cells, reducing or suppressing an immune response, inducing angiogenesis,
 CC correcting muscle contractility, cardiac hypertrophy, cardiac
 CC insufficiency and myocarditis. (Iib) is useful for expressing a gene of
 CC therapeutic interest in vivo, by isolating (Iib) and introducing (Iib) in
 CC the cardiac tissue, under conditions so that the gene of interest is
 CC expressed. (I), the vectors and the compositions are useful in clinical,
 CC experimental, therapeutic and diagnostic fields, and in the treatment and
 CC prevention of cardiac pathologies. (I) is also useful for generating
 CC transgenic animals which constitute models for studying certain cardiac
 CC pathologies. The transgenic animals are also useful for screening
 CC molecules for their activity on the regulatory sequences of the gene
 CC encoding the CARP protein. The present sequence represents the DNA
 CC fragment upstream of the coding sequence of a mouse CARP protein.
 XX
 SQ Sequence 2358 BP; 636 A; 567 C; 542 G; 613 T; 0 other;

Query Match 17.7%; Score 367.8; DB 24; Length 2358;
 Best Local Similarity 76.4%; Pred. No. 7.3e-76;
 Matches 597; Conservative 0; Mismatches 162; Indels 22; Gaps 11;
 Qy 1289 TCTATTTCTTGACCACTCTGATCCATTTTGAAGTAAAAATGCTCCAATTTATTTGCTGTT 1348
 Db 1540 TCTCTGTGTCATCACTTCGGCCCGTTTGGGGT--AGATCTCTGATTAGCCTTCAGAT 1597
 Qy 1349 TTAGAACACGGTAAGCATGTCATGTGCTA---ATGCCAGTGCATCATAAAGAAAGT 1405
 Db 1598 TTAGAACACGGTCAAGCCTGTGTGCATTAATATGGCCAGTGCACCATAGTCAAGT 1657
 Qy 1406 GCATTACTGAATGCTTTCATGCTTATTAATGATGTAAGTGGCATGTATGGGCGCTA 1465
 Db 1658 GCATTACTGAATGCTTTCATGCTTCTCTAATGCTGTGATGCGATGTCACAGGGCCAT 1717
 Qy 1466 TTTAGC-CCAGACATCACTCCAAAGAAATTCCAAACAGATATAGACAAGTGCCTTTAGGGC 1524
 Db 1718 TTTAGTGCAGACATCACTCCAGAGAAATTCCAAACAGATAGAGACAAGTGGCACCCAGAC 1777
 Qy 1525 CCAGATCCCTTCCCTCAGGCTGTTTACCCAGGGAATAGGATGTCC---TGGGACAAGTT 1581
 Db 1778 CCATCTCTCTCCCTCGGGCTGATTTATCCCAAGAAATAGGATGTCCCAAGCAACACTTC 1837
 Qy 1582 TCCCTTAAGTGAAGTGTGATAGTCTGCTTATCAGAAAGATATTACTGGGGTGTGATA 1641
 Db 1838 CCAGCCAACTGGAGTGTGATAGTCCAGTTATCAGAAAGATATGCTGTAAGTGTGATG 1897
 Qy 1642 TGTAGGGCATCTACATTTTCTTTGATA-GGTAGTTCATATCAAAGCTGACAAAGAA--AAAA 1698
 Db 1898 CACAGTGC--TTGCATTTTCTTGATACGTTAGTCATATGAGAGCTGACAAAGAGGAAA 1955
 Qy 1699 AGGGCAGTGAATGTTGCAATGTCACAGACAGCTGTCCCTGAC-TCTTGACAAATAG 1757
 Db 1956 AGAGCAGCGATGTTGTCATTAATTAACAGGAGCTGTCCCTGGCTTCCGATACGTGGG 2015
 Qy 1758 ATGACTTGCAATGCTGAGCGATGTCATCACCACCAAGGAATGGCCCTCTCACATTTCTT 1817
 Db 2016 ATGACTCGCAATGCTGAGCGGTGTGGTCACTGCCAAAGGAATGACCTCTCACATTTCTT 2075
 Qy 1818 CTTGATTCACATATTTTCCAGCAGGGTTAGCTTGTCTCCCTCCCTCTTCAGCTTCCAGAC 1877
 Db 2076 CTTGATTCGATACCGCGGG-----CCAGCTTGTGATCTCTCTCTTGGGCTTCCAGAC 2130
 Qy 1878 ACTGAGTCTGGAATGAAAAATTCACCTGCTCTGAGTGTGCTCTAATATGGGGCGGAGTG 1937
 Db 2131 ACTAAGTCTGGAATGAAAAATTCACCTGCTCTGAAATTTGCCACTGTGTGGGCGAGGGTG 2190
 Qy 1938 TTACTTGGTTCCTCCAGGTTGGAAGATTTATCTACCGGGCCCCAGCTATATAGCTACCG 1997
 Db 2191 TGACTTGGCTTCCAGGCTGGAAGATTTATCTACCGAGCCCTAGCTATATAA-CGGGCTG 2249

QY 1998 GTGTGGGGGCCAGCAGGGCCAACTCCAGGATTCCTTC-CACGACAGAAAAACATAC 2056
 |||||
 Db 2250 GTGTGGGGGCTCCACAGGGCCAGTTCAGGGGTTTCATCCACAGAGAGAAAAACATAG 2309
 |||||
 QY 2057 A 2057
 |||||
 Db 2310 A 2310
 |||||

RESULT 3
 AAA10406
 ID AAA10406 standard; DNA; 2247 BP.
 AC AAA10406;
 XX
 DT 18-JUL-2000 (first entry)
 XX Murine cardiac ankyrin repeat protein (CARP) promoter.
 DE
 XX
 XX Cardiac ankyrin repeat protein; CARP promoter; murine; adenovirus vector;
 KW cardiac specific; heart disease; gene therapy; ds.
 XX
 OS Mus musculus.
 XX
 PN WO200015821-A1.
 XX
 PD 23-MAR-2000.
 XX
 XX 10-SEP-1999; 99WO-US20730.
 PF
 XX 11-SEP-1998; 98US-0099960.
 PR
 XX (REGC) UNIV CALIFORNIA.
 PA
 XX Chien KR, Wang Y, Evans S;
 PI WPI; 2000-271457/23.
 DR
 XX Human type-5 recombinant adenovirus vector used for targeted gene
 PT therapy for heart disease and evaluating gene function contains a
 PT tissue-restricted promoter and inverted terminal repeat sequences -
 XX
 PS Claim 8; Page 29-30; 33pp; English.
 XX
 CC The invention relates to a human type-5 recombinant adenovirus vector
 CC for achieving cardiac-restricted transcription of a gene of interest.
 CC The vector comprises inverted terminal repeat (ITR) sequences from human
 CC adeno-associated virus (AAV) type 2 (AAA10404-A10405) and a cardiac
 CC tissue-specific promoter. In particular, the promoter is that of the
 CC cardiomyocyte-restricted cardiac ankyrin repeat protein (CARP) gene.
 CC The adenovirus vector is used for targeted gene therapy for heart
 CC disease and for evaluating gene function. Cardiac restricted
 CC transcription of a transgene in both neonatal and mature cardiac tissues
 CC can be achieved to treat inherited and acquired heart diseases. The
 CC vector is suitable for tissue-specific use in vivo and in vitro and
 CC provides cardiac restricted transcription. The present sequence
 CC represents the murine cardiac ankyrin repeat protein (CARP) promoter.
 XX
 SQ Sequence 2247 BP; 609 A; 549 C; 485 G; 589 T; 15 other;

Query Match 15.6%; Score 322.6; DB 21; Length 2247;
 Best Local Similarity 76.3%; Pred. No. 2.5e-65;
 Matches 567; Conservative 0; Mismatches 154; Indels 22; Gaps 13;

QY 1325 AAATGCTCCAATTATTATGCTGTTTGTAGAACACGGTAAGCATGTCAT--GTGCTAATGGC 1382
 |||||
 Db 1499 AGATCTCTCATTTAGGCTTCAGATTTTGAACACGGTGGCTGTGCTACTAATTAATGGC 1558
 |||||
 QY 1383 CAGTGACATCAAAAAGAAAGTGCATTTACTGAATGCTTTCAATGCTTTATATATGATGGT 1442
 |||||
 Db 1559 CAGTGACACCATAGAGTCAAAAGTGCATTTACTGAATGCTTTCAATTTCTCTAATGCTGGT 1618
 |||||
 QY 1443 AAGGTGGCATGTCATGGGGCTATTAGCCGACATCATCTCCAAAGAATTCCAAACAGA 1502
 |||||

Db 1619 AGATGGCATGTCCAGGGCCATTTTAGCTGCAGACATCATCCAGAGAAATTCCAAACAGA 1678
 |||||
 QY 1503 TATAGCAAGTGCCTTTTAGGGCCAGATCCCTTCCCTCAGGCTGTTTACCAGGGAATA 1562
 |||||
 Db 1679 TA-GGACAAGTGGCACCAGACCCATCT-CTTCCCTCGGCTGATTATCCCAAATA 1736
 |||||
 QY 1563 GGATGTCC---TGGGACAAGTTTCCCTTAAGTGAAGTGTGATAAGTCTGCTTATCAGAA 1619
 |||||
 Db 1737 GGATGTCCCAAGCAACACTTCCAGCCAACTGGAGTGTGATAAGTCCAGTTATCAGAA 1796
 |||||
 QY 1620 AGATATTACTGGGGTGTGATATGTAGGCATCTACATTTCTTTGATA-GGTAGTCAAT 1678
 |||||
 Db 1797 AGATATGGCTGTAAAGTGTGATGCACAGTGC--TTGCAATTTCTTGATAGTTAGTCAAT 1854
 |||||
 QY 1679 GAAAGCTGACAAAGAA--AAAAAGGCGCAGTGTGTCGCAATGTCAACAGACAGCTGTC 1736
 |||||
 Db 1855 GAGAGCTGACAAAGAAAGGAAAGAGAGCAGCGATGT-GTGCAATATTAAACAGGCAGCTGTC 1913
 |||||
 QY 1737 CCTGTAC-TCTTGACAAATAGGATGACTTGCATTTGCTGAGCGATGTGATCACCACCAAG 1795
 |||||
 Db 1914 CCTGGCTTCCGATACGTTGGGATGACTCGCATTTGCTGAGCGGTGTGGTCACTGCCAAAG 1973
 |||||
 QY 1796 GAATGSCCTCTCACATTTCTTCTCGATTTCATATTTCAGCAGGGTGTAGTCTGCTCTCCC 1855
 |||||
 Db 1974 GAATGACCTCTCACATTTCTTCTCGATTTCGATTCGATACGCCGCGG-----CGAGTTGTCTAT 2028
 |||||
 QY 1856 CTCCTCTTTTTCAGCTTCCAGACACTGAGTCTGGAATGAAATTCACCTGCTCTGAGTTG 1915
 |||||
 Db 2029 CTCCTCTTTGGGCTTCCAGACACTAAGTCTGGAATGAAATTCACCTGCTCTGAAATTG 2088
 |||||
 QY 1916 GCTCTTAATGGGGGGGAGTGTACTTCGGTTCCAGGTTGGAGATTATCTCACCCGG 1975
 |||||
 Db 2089 GCCACTGTGGGAGCAGGGGTGTGACTTGGCTTCCAGGCTGGAAGATTATCTCACCCAG 2148
 |||||
 QY 1976 CCCAGCTATTAAGCTGACCGGTGTGGAGGGGCCAGCAGGGCCAACTCCAGGGAATTC 2035
 |||||
 Db 2149 CCTA-CTATATAA-CGGGCTGTGTGGAGGGGCTCCAGAGGCCAGTTCAGGGGTTCA 2206
 |||||

RESULT 4
 AA193508
 ID AA193508 standard; cDNA; 1988 BP.
 AC AA193508;
 XX
 DT 06-NOV-2001 (first entry)
 XX Human polynucleotide SEQ ID NO 13568.
 DE
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KW nervous system disorders; arthritis; inflammation; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200164835-A2.
 XX
 PD 07-SEP-2001.
 XX
 XX 26-FEB-2001; 2001WO-US04927.
 XX
 PR 28-FEB-2000; 2000US-0515126.
 PR 18-MAY-2000; 2000US-0577409.
 XX
 PA (HYSE-) HYSEQ INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;

XX DR WPI; 2001-514838/56.
XX DR P-PSDB; AAO13577.
XX
XX Isolated nucleic acids and polypeptides, useful for preventing
XX PT diagnosing and treating e.g. leukaemia, inflammation and immune
XX PT disorders -
XX
XX Claim 1; SEQ ID NO 13568; 1399pp + Sequence Listing; English.
XX
XX The invention relates to human polynucleotides (AA179941-AA193841) and
XX CC the encoded proteins (AA000010-AA013910) that exhibit activity elating to
XX CC cytokine, cell proliferation or cell differentiation or which may induce
XX CC production of other cytokines in other cell populations. The
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
XX CC peptide therapy. The polypeptides have various cytokine-like activities,
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating
XX CC activity, tissue growth factor activity, immunomodulatory activity and
XX CC activin/inhibin activity and may be useful in the diagnosis and/or
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
XX CC inflammation.
XX CC Note: The sequence data for this patent did not form part of the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 1988 BP; 632 A; 384 C; 472 G; 500 T; 0 other;
XX
Query Match 11.1%; Score 229.6; DB 22; Length 1988;
Best Local Similarity 98.0%; Pred. No. 1.2e-43;
Matches 243; Conservative 0; Mismatches 4; Indels 1; Gaps 1;
XX
QY 1828 ATATTCAGCAGGGTTAGTCTGT-CCTCCCTCCCTCTTCAGCTTCCGACACTGAGTCT 1886
Db 2 AAAAAACAGCAGGGTTAGTCTGTCCCTCCCTCTTCAGCTTCCGACACTGATCT 61
XX
QY 1887 GGAATGAAATTCACCTGCCTCTGAGTTGGCTCTAATGGGCGGGAGTGTACTTCGG 1946
Db 62 GGAATGAAATTCACCTGCCTCTGAGTTGGCTCTAATGGGCGGGAGTGTACTTCGG 121
XX
QY 1947 TTCCAGGTTGGAAGATTATCTCACCGGCCCGAGCTATATAGCTGACCGGTGTGGAGG 2006
Db 122 TTCCAGGTTGGAAGATTATCTCACCGGCCCGAGCTATATAGCTGACCGGTGTGGAGG 181
XX
QY 2007 GGCCACGAGGGCCAACTCCAGGGATTCTTCACGACAGAAAAACATACAGACTCCTT 2066
Db 182 GGCCACGAGGGCCAACTCCAGGGATTCTTCACGACAGAAAAACATACAGACTCCTT 241
XX
QY 2067 CAGCCCAAC 2074
Db 242 CAGCCCAAC 249
XX
RESULT 5
AAH02910
ID AAH02910 standard; DNA; 1901 BP.
XX
XX AAH02910;
XX
XX
XX 15-JUN-2001 (first entry)
XX
XX DE Human shear stress-response coding sequence SEQ ID NO: 73.
XX
XX Human; shear stress-response protein; vascular disease;
XX KW arteriosclerosis; ds.
XX
XX OS Homo sapiens.
XX
XX PN WO200125427-A1.
XX
XX
XX PD 12-APR-2001.
XX
XX PF 02-OCT-2000; 2000WO-JP06840.
XX
XX

PR 01-OCT-1999; 99JP-0280976.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
XX PA (NOJI/) NOJIMA H.
XX
XX Nojima H, Yoshisue H, Obayashi M, Ota T, Kawabata A, Sakurada K;
XX PI Kuga T, Sekine S, Nakamura Y, Sugano S;
XX
XX WPI; 2001-266308/27.
XX DR P-PSDB; AAB90787.
XX
XX DNA sequences, proteins encoded by them and antibodies against them
XX PT useful in diagnosis and treatment of vascular disease caused by
XX PT arteriosclerosis -
XX
XX Claim 20; Page 422-425; 678pp; Japanese.
XX
XX The present invention provides the protein and coding sequences of a
XX CC number of human shear stress response proteins. These are useful in the
XX CC diagnosis, treatment and screening of vascular diseases caused by
XX CC arteriosclerosis, including heart failure, post-PTCA restenosis and
XX CC hypertension.
XX
XX Sequence 1901 BP; 592 A; 378 C; 460 G; 471 T; 0 other;
XX
Query Match 11.0%; Score 228; DB 22; Length 1901;
Best Local Similarity 97.6%; Pred. No. 2.8e-43;
Matches 242; Conservative 0; Mismatches 5; Indels 1; Gaps 1;
XX
QY 1828 ATATTCAGCAGGGTTAGTCTGT-CCTCCCTCCCTCTTCAGCTTCCGACACTGAGTCT 1886
Db 2 AAAAAACAGCAGGGTTAGTCTGTCCCTCCCTCTTCAGCTTCCGACACTGATCT 61
XX
QY 1887 GGAATGAAATTCACCTGCCTCTGAGTTGGCTCTAATGGGCGGGAGTGTACTTCGG 1946
Db 62 GGAATGAAATTCACCTGCCTCTGAGTTGGCTCTAATGGGCGGGAGTGTACTTCGG 121
XX
QY 1947 TTCCAGGTTGGAAGATTATCTCACCGGCCCGAGCTATATAGCTGACCGGTGTGGAGG 2006
Db 122 TTCCAGGTTGGAAGATTATCTCACCGGCCCGAGCTATATAGCTGACCGGTGTGGAGG 181
XX
QY 2007 GGCCACGAGGGCCAACTCCAGGGATTCTTCACGACAGAAAAACATACAGACTCCTT 2066
Db 182 GGCCACGAGGGCCAACTCCAGGGATTCTTCACGACAGAAAAACATACAGACTCCTT 241
XX
QY 2067 CAGCCCAAC 2074
Db 242 CAGCCCAAC 249
XX
RESULT 6
AAD27217
ID AAD27217 standard; DNA; 1901 BP.
XX
XX AAD27217;
XX
XX
XX 09-APR-2002 (first entry)
XX
XX DE Human CAA58676 protein encoding EST clone X83703 DNA.
XX
XX Human; congestive heart failure; dilative cardiomyopathy; sudden death;
XX KW hypertrophic cardiomyopathy; ischaemic cardiomyopathy; rhythm disorder;
XX KW heart muscle disease; conduction disorder; coronary heart disease;
XX KW systemic arterial hypertension; pulmonary hypertension; endocarditis;
XX KW pulmonary heart disease; valvular heart disease; pericardial disease;
XX KW congenital heart disease; gene therapy; syncope; transgenic animal;
XX KW expressed sequence tag; EST; clone X83703; CAA58676 protein; ds.
XX
XX OS Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX FH 250..1209
XX CDS /*tag= a
XX
XX FT

```
FT misc_feature /product= "Human CAA58676 protein"
FT 198..431
FT /*tag= b
FT /note= "6268 cDNA fragment"
FT misc_feature
FT 1604..1753
FT /*tag= b
FT /note= "S1MC01-1 cDNA fragment"
FT misc_feature
FT 1302..1306
FT /*tag= b
FT /note= "AU-rich mRNA decay element"
FT misc_feature
FT 1391..1401
FT /*tag= b
FT /note= "AU-rich mRNA decay element"
FT misc_feature
FT 1415..1423
FT /*tag= b
FT /note= "AU-rich mRNA decay element"
FT
FT
PN WO200192567-A2.
XX
XX
XX 06-DEC-2001.
XX
XX 30-MAY-2001; 2001WO-EP06165.
XX
XX 30-MAY-2000; 2000US-207400P.
XX
XX (MEDI-) MEDIGENE AG.
XX
XX Bunk D, Reuner B, Beck J, Henkel T;
XX
XX WPI; 2002-122073/16.
XX P-PSDB; AAE16633.
XX
XX Identifying a subject at risk for a heart disease e.g. congestive heart
XX failure, dilative cardiomyopathy, heart muscle disease, by quantifying
XX the polypeptide expressed by genes abnormally expressed in heart tissue
XX
XX
XX Claim 2a; Fig 10b; 154pp; English.
XX
XX The patent discloses novel target genes abnormally expressed in heart
XX tissues and their corresponding proteins. The invention also relates to
XX methods for assessing the expression level of these genes. The method
XX is used for testing the predisposition of mammals and preferably humans
XX for a heart disease or for an acute state of such a disease. It is also
XX useful to treat diseases of the heart such as congestive heart failure,
XX dilative cardiomyopathy, hypertrophic cardiomyopathy, ischaemic cardio-
XX myopathy, specific heart muscle disease, rhythm and conduction disorders,
XX syncope and sudden death, coronary heart disease, systemic arterial
XX hypertension, pulmonary hypertension, pulmonary heart disease, valvular
XX heart disease, congenital heart disease, pericardial disease and
XX endocarditis. Sequences of the invention are also used in gene therapy.
XX A transgenic non-human mammal comprising the sequences of the invention
XX are useful for the development for medicaments for the treatments of
XX heart diseases. The present DNA sequence is expressed sequence tag
XX (EST) clone X83703 which encodes CAA58676 protein.
XX
XX Sequence 1901 BP; 592 A; 378 C; 460 G; 471 T; 0 other;
XX
XX
XX Query Match 11.0%; Score 228; DB 24; Length 1901;
XX Best Local Similarity 97.6%; Pred. No. 2.8e-43;
XX Matches 242; Conservative 0; Mismatches 5; Indels 1; Gaps 1;
XX
XX 1828 ATATTGACGAGGTTAGTTGT-CTTCCCTCCCTCTTACGTTTCCGACACTGATCT 1886
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 2 AAAAAGACGAGGTTAGTTGTGTCCTCCCTCCCTCTTACGTTTCCGACACTGATCT 61
XX
XX 1887 GGAATGAAATTCACCTGCTCTGAGTTGGCTCCTTAATGGGGCGGAGTGTACTTCGG 1946
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 62 GGAATGAAATTCACCTGCTCTGAGTTGGCTCCTTAATGGGGCGGAGTGTACTTCGG 121
XX
XX 1947 TTCCCGAGTTGGAAGATTATCTCACCCGGCCCGGCTATATAAGCTGACCGGTTGGAGG 2006
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 122 TTCCCGAGTTGGAAGATTATCTCACCCGGCCCGGCTATATAAGCTGACCGGTTGGAGG 181
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 2007 GGCCGAGCAGGCCCAACTCCAGGATTCCTTCCAGCAGAGAAAACATACAGACTCCTT 2066
DB 182 GGCCGAGCAGGCCCAACTCCAGGATTCCTTCCAGCAGAGAAAACATACAGACTCCTT 241
QY 2067 CAGCCAAC 2074
DB 242 CAGCCAAC 249

RESULT 7
AAS94903
ID AAS94903 standard; DNA; 1889 BP.
XX
XX AAS94903;
AC
XX
XX 14-FEB-2002 (first entry)
XX
XX Human DNA sequence #158 expressed during foam cell differentiation.
DE
XX
XX Human; foam cell differentiation; atherosclerosis; cerebral stroke;
KW
XX cardiovascular disorder; coronary artery disease; gene therapy; ds.
XX
XX Homo sapiens.
OS
XX
XX WO200177389-A2.
PN
XX
XX 18-OCT-2001.
PD
XX
XX 04-APR-2001; 2001WO-US111128.
PF
XX
XX 05-APR-2000; 2000US-195106P.
PR
XX
XX (INCY-) INCYTE GENOMICS INC.
PA
XX
XX Shiffman D, Somogyi R, Lawn R, Seilhamer JJ, Porter GU, Mikita T;
PI
XX Tai J;
PI
XX WPI; 2002-010925/01.
XX
XX
XX Composition useful for diagnosis of conditions, disorders or diseases
XX associated with atherosclerosis, comprises several polynucleotides that
XX are differentially expressed in foam cell development -
XX
XX Claim 1; Page 213; 315pp; English.
XX
XX The present invention relates to the isolation of human polynucleotide
XX sequences that are differentially expressed during foam cell
XX differentiation. The polynucleotide sequences of the invention or a
XX composition comprising these polynucleotides are useful as a high
XX throughput method for detecting altered expression of one or more
XX polynucleotides in a sample. The polynucleotides can be used in the
XX diagnosis of disorders associated with foam cell development such as
XX atherosclerosis, cerebral stroke, and cardiovascular disorders such as
XX coronary artery disease. The polynucleotide sequences can also be used
XX as PCR primers and probes. The polynucleotides of the invention are also
XX useful in gene therapy. AAS94746-AAS95021 represent the human
XX polynucleotide sequences of the invention which are differentially
XX expressed during foam cell differentiation.
XX
XX Sequence 1889 BP; 613 A; 354 C; 452 G; 442 T; 28 other;
XX
XX
XX Query Match 7.0%; Score 144.4; DB 24; Length 1889;
XX Best Local Similarity 99.3%; Pred. No. 8.9e-24;
XX Matches 145; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX 1929 GCGGGAGTCTTACTTCGGTTCCAGGTTGGAAGATTATCTCACCCGGCCCGGCTATATA 1988
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 5 GAGGGAGTCTTACTTCGGTTCCAGGTTGGAAGATTATCTCACCCGGCCCGGCTATATA 64
XX
XX 1989 AGCTGACCGGTGTGGAGGGGCCAGCAGGGCCCAACTCCAGGATTCCTTCCAGCAGAA 2048
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
XX 65 AGCTGACCGGTGTGGAGGGGCCAGCAGGGCCCAACTCCAGGATTCCTTCCAGCAGAA 124
XX | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```


QY 2049 AACATACAAGACTCCTTCAGCCAAC 2074
DB 125 AACATACAAGACTCCTTCAGCCAAC 150

RESULT 8
ID ABX77648 standard; cDNA; 1889 BP.
AC ABX77648;
XX
XX
DT 09-APR-2003 (first entry)
XX
DE Differentially expressed breast cancer associated cDNA #143.
XX
KW Breast cancer; differential gene expression; BC-cDNA;
KW breast cancer diagnosis; breast cancer monitoring;
KW breast cancer treatment; breast cancer staging; gene; ss.
XX
OS Homo sapiens.
XX
XX
PN US2002156263-A1.
XX
PD 24-OCT-2002.
XX
PF 04-OCT-2001; 2001US-0974298.
XX
PR 05-OCT-2000; 2000US-238331P.
XX
PA (CHEN/) CHEN H.
PI Chen H;
XX
PI WPI; 2003-182653/18.
XX
XX
PT New cDNAs, which are differentially expressed in (metastatic) breast
PT cancer useful for diagnosing or staging, breast cancer, or for
PT monitoring the treatment of breast cancer in an individual -
XX
XX
PS Claim 1; SEQ ID NO 182; 30pp; English.
XX
XX
CC The invention describes a combination of cDNAs (designated BC-cDNAs),
CC which are differentially expressed in breast cancer. The combination
CC includes 152 cDNA sequences, or their complements. The protein encoded
CC by any of these BC-cDNAs is useful for screening several molecules or
CC compounds to identify at least one ligand that specifically binds the
CC protein, producing or preparing polyclonal or monoclonal antibodies, or
CC purifying antibodies from a sample. The antibodies, which specifically
CC bind the protein differentially expressed in breast cancer is useful for
CC detecting the expression of a protein in a sample. The BC-cDNAs are
CC also useful for diagnosing, monitoring the treatment of, or staging,
CC breast cancer. This sequence represents a differentially expressed
CC breast cancer associated cDNA.
CC Note: The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from the US patent office at
CC seqdata.uspto.gov/sequence.html?DocID=20020156263.
XX
XX
SQ Sequence 1889 BP; 613 A; 354 C; 452 G; 469 T; 1 other;
Query Match 7.0%; Score 144.4; DB 25; Length 1889;
Best Local Similarity 99.3%; Pred. No. 8.9e-24;
Matches 145; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1929 GCGGAGTGTTACTTCGGTTCCAGGTTGGAAGATTATCTCACCGGCCCCAGCTATATA 1988
DB 5 GAGGAGTGTTACTTCGGTTCCAGGTTGGAAGATTATCTCACCGGCCCCAGCTATATA 64
QY 1989 AGCTACCGGTTGAGGGGGCCCCAGGAGGCAACTCCAGGAGTTCCTTCCAGCAGNA 2048
DB 65 AGCTACCGGTTGAGGGGGCCCCAGGAGGCAACTCCAGGAGTTCCTTCCAGCAGNA 124

QY 2049 AACATACAAGACTCCTTCAGCCAAC 2074
DB 125 AACATACAAGACTCCTTCAGCCAAC 150

RESULT 9
ID AAF97854 standard; DNA; 34488 BP.
XX
AC AAF97854;
XX
XX
DT 31-MAY-2001 (first entry)
XX
DE Human neuroblastoma cell line NB-1 lp36 nucleotide sequence SEQ ID NO:68.
XX
KW Human; chromosome 1; lp36; neuroblastoma cell line; NB-1; anticancer;
KW tumour suppressor; human lp36 homozygosity deletion domain; tumour;
KW diagnosis; ds.
XX
OS Homo sapiens.
XX
XX
PN WO200116311-A1.
XX
PD 08-MAR-2001.
XX
XX
PF 31-AUG-2000; 2000WO-JP05930.
XX
PR 31-AUG-1999; 99JP-0245962.
PR 09-MAY-2000; 2000JP-0136266.
XX
XX (HISM) HISAMITSU PHARM CO LTD.
PA (CHIB-) CHIBA PREFECTURE.
PI Nakagawara A;
XX
XX
DR WPI; 2001-226686/23.
XX
XX
PT Human lp36 homozygosity deletion domain from the 36-position of first
PT chromosome short arm in human neuroblastoma cell lines, applicable e.g.
PT in gene diagnosis of tumors as well as in developing anti-cancer drugs
PT
XX
XX
PS Example 8; Page 104-118; 226pp; Japanese.
XX
XX
CC The present invention describes a homozygosity deletion domain
CC co-existing in the 36-position of the first chromosome short arm (lp36)
CC in human neuroblastoma. Also described are base sequences from the lp36
CC position of human neuroblastoma cell lines (NB-1 and MASS-NB-SCH-1),
CC which are tumour suppressor genes in human neuroblastoma. The genes are
CC tumour suppressor genes, base sequence data of which are applicable as
CC tumour markers and reagents in studying mechanism of tumour body
CC formation, and gene diagnosis of tumours as well as in developing
CC anti-cancer drugs. AAF9787 to AAF97829 represent PCR primers used in
CC the exemplification of the present invention, and AAF97830 to AAF97874
CC represent sequences given in the exemplification of the present
CC invention.
XX
XX
SQ Sequence 34488 BP; 9654 A; 6717 C; 6926 G; 11191 T; 0 other;
Query Match 4.7%; Score 97.2; DB 22; Length 34488;
Best Local Similarity 74.1%; Pred. No. 2.4e-12;
Matches 123; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY 258 AAATAGTAATGCTCTGTTGCTTCAGCAGGCATATATAAATAGGAGCTATACAAG 317
DB 20646 AATAACTAAGCAGCGTGCTCGCTTCGGCAGCAGCATATATAAATTTGGAACCATACAG 20705
QY 318 AAGATTAGCATGCATCTGTGCAAGAATGACACAAATTTGTGAAACATTCATATATT 377
DB 20706 AAGATTAGCATGGCCCTTCGGCAAGATGACACGCAAAATTCGTGAAGCGTTCATATTTA 20765
QY 378 AAAAAATAAATAATAAAGAGAAAGAAAAATTAAGAAAA 423

Db 20766 AAAAAAAAAAAGAAACAAACAAACAAACAAACAAACAAACAAACAAACAA 20811

RESULT 10

AAK68280
ID AAC66548 standard; DNA; 121162 BP.

XX AC AAC66548;

XX DT 19-FEB-2001 (first entry)

XX XX Human kinesin-like protein HKLP coding sequence contig SEQ ID NO: 1.

XX XX Human; kinesin-like protein; HKLP; KIF1; cell division; cancer;

XX KW intracellular transport; neurological disorder; infertility; ds.

XX KW biallelic marker; spontaneous abortion; neonatal chromosome disorder;

XX KW aneuploidy; ds.

XX OS Homo sapiens.

XX PN WO200063375-A1.

XX PD 26-OCT-2000.

XX PF 20-APR-2000; 2000WO-IB00562.

XX PP 20-APR-1999; 99US-0130217.

XX PA (GEST) GENSET.

XX PI Bougueleret L, Dufaure-Gare I, Grel P;

XX DR WPI; 2000-665242/64.

XX PT An isolated or purified human kinesin-like protein (HKLP) encoding

XX PT polynucleotide used to detect HKLP polynucleotides in a sample

XX PT comprises a contiguous span of at least 12 nucleotides -

XX PS Claim 1; Page 143-175; 199pp; English.

XX CC The present invention describes the coding and protein sequences of the

XX CC human kinesin-like protein HKLP. It is thought that the protein could be

XX CC involved in neurological disorders, infertility, spontaneous abortion,

XX CC neonatal chromosome disorders, aneuploidy and cancers. This is due to its

XX CC function in the movement of microtubules. The protein shows homology to

XX CC the murine KIF1A and KIF1B proteins. The sequences disclosed in the

XX CC invention can be used in the isolation of similar human proteins and in

XX CC vector production. In addition, the biallelic markers shown can be used

XX CC in disease diagnosis and population studies.

XX SQ Sequence 121162 BP; 33272 A; 24108 C; 25842 G; 37919 T; 21 other;

Query Match 4.7%; Score 97.2; DB 21; Length 121162;

Best Local Similarity 74.1%; Pred. No. 3.6e-12;

Matches 123; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

OY 258 AATAGTATGACTCTGTTGCTTTCAGCAGGACATATCTAAATAGGAGCTATACAAG 317

DB 47492 AATAACATAGCAGCGTGCTCGTTCGGCAGCATATCTAAATTTGGAACGATACAGAG 47551

OY 318 AAGATTAGCATGGACTCTGTGCAAGAAATGACACACAAATTTGGAACATTTCCATATATT 377

DB 47552 AAGATTAGCATGGCCCTCGGCAAGATGACGCAAAATTCGGAAGCTTCCATATTTA 47611

AAK68280;
06-NOV-2001 (first entry)
Human immune/haematopoietic antigen genomic sequence SEQ ID NO:23092.
Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
cytostatic; gene therapy; vaccine; metastasis; ds.

OS Homo sapiens.

PN WO200157182-A2.

XX 09-AUG-2001.

XX 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

XX 04-FEB-2000; 2000US-0180628.

XX 24-FEB-2000; 2000US-0184664.

XX 02-MAR-2000; 2000US-0186350.

XX 16-MAR-2000; 2000US-0189874.

XX 17-MAR-2000; 2000US-0190076.

XX 18-APR-2000; 2000US-0198123.

XX 19-MAY-2000; 2000US-0205515.

XX 07-JUN-2000; 2000US-0209467.

XX 28-JUN-2000; 2000US-0214886.

XX 30-JUN-2000; 2000US-0215135.

XX 07-JUL-2000; 2000US-0216647.

XX 07-JUL-2000; 2000US-0216880.

XX 11-JUL-2000; 2000US-0217487.

XX 14-JUL-2000; 2000US-0217496.

XX 26-JUL-2000; 2000US-0218290.

XX 26-JUL-2000; 2000US-0220963.

XX 14-AUG-2000; 2000US-0220964.

XX 14-AUG-2000; 2000US-0224518.

XX 14-AUG-2000; 2000US-0224519.

XX 14-AUG-2000; 2000US-0225213.

XX 14-AUG-2000; 2000US-0225214.

XX 14-AUG-2000; 2000US-0225266.

XX 14-AUG-2000; 2000US-0225267.

XX 14-AUG-2000; 2000US-0225268.

XX 14-AUG-2000; 2000US-0225270.

XX 14-AUG-2000; 2000US-0225447.

XX 14-AUG-2000; 2000US-0225757.

XX 14-AUG-2000; 2000US-0225758.

XX 18-AUG-2000; 2000US-0226279.

XX 22-AUG-2000; 2000US-0226681.

XX SQ

RESULT 11

AAK68280
ID AAK68280 standard; DNA; 7141 BP.

XX AC AAK68280;

XX DT 19-FEB-2001 (first entry)

XX XX Human kinesin-like protein HKLP coding sequence contig SEQ ID NO: 1.

XX XX Human; kinesin-like protein; HKLP; KIF1; cell division; cancer;

XX KW intracellular transport; neurological disorder; infertility; ds.

XX KW biallelic marker; spontaneous abortion; neonatal chromosome disorder;

XX KW aneuploidy; ds.

XX OS Homo sapiens.

XX PN WO200063375-A1.

XX PD 26-OCT-2000.

XX PF 20-APR-2000; 2000WO-IB00562.

XX PP 20-APR-1999; 99US-0130217.

XX PA (GEST) GENSET.

XX PI Bougueleret L, Dufaure-Gare I, Grel P;

XX DR WPI; 2000-665242/64.

XX PT An isolated or purified human kinesin-like protein (HKLP) encoding

XX PT polynucleotide used to detect HKLP polynucleotides in a sample

XX PT comprises a contiguous span of at least 12 nucleotides -

XX PS Claim 1; Page 143-175; 199pp; English.

XX CC The present invention describes the coding and protein sequences of the

XX CC human kinesin-like protein HKLP. It is thought that the protein could be

XX CC involved in neurological disorders, infertility, spontaneous abortion,

XX CC neonatal chromosome disorders, aneuploidy and cancers. This is due to its

XX CC function in the movement of microtubules. The protein shows homology to

XX CC the murine KIF1A and KIF1B proteins. The sequences disclosed in the

XX CC invention can be used in the isolation of similar human proteins and in

XX CC vector production. In addition, the biallelic markers shown can be used

XX CC in disease diagnosis and population studies.

XX SQ Sequence 121162 BP; 33272 A; 24108 C; 25842 G; 37919 T; 21 other;

Query Match 4.7%; Score 97.2; DB 21; Length 121162;

Best Local Similarity 74.1%; Pred. No. 3.6e-12;

Matches 123; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

OY 258 AATAGTATGACTCTGTTGCTTTCAGCAGGACATATCTAAATAGGAGCTATACAAG 317

DB 47492 AATAACATAGCAGCGTGCTCGTTCGGCAGCATATCTAAATTTGGAACGATACAGAG 47551

OY 318 AAGATTAGCATGGACTCTGTGCAAGAAATGACACACAAATTTGGAACATTTCCATATATT 377

DB 47552 AAGATTAGCATGGCCCTCGGCAAGATGACGCAAAATTCGGAAGCTTCCATATTTA 47611

XX SQ

RESULT 11

AAK68280
ID AAK68280 standard; DNA; 7141 BP.

XX AC AAK68280;

XX DT 19-FEB-2001 (first entry)

XX XX Human kinesin-like protein HKLP coding sequence contig SEQ ID NO: 1.

XX XX Human; kinesin-like protein; HKLP; KIF1; cell division; cancer;

XX KW intracellular transport; neurological disorder; infertility; ds.

XX KW biallelic marker; spontaneous abortion; neonatal chromosome disorder;

XX KW aneuploidy; ds.

XX OS Homo sapiens.

XX PN WO200063375-A1.

XX PD 26-OCT-2000.

XX PF 20-APR-2000; 2000WO-IB00562.

XX PP 20-APR-1999; 99US-0130217.

XX PA (GEST) GENSET.

XX PI Bougueleret L, Dufaure-Gare I, Grel P;

XX DR WPI; 2000-665242/64.

XX PT An isolated or purified human kinesin-like protein (HKLP) encoding

XX PT polynucleotide used to detect HKLP polynucleotides in a sample

XX PT comprises a contiguous span of at least 12 nucleotides -

XX PS Claim 1; Page 143-175; 199pp; English.

XX CC The present invention describes the coding and protein sequences of the

XX CC human kinesin-like protein HKLP. It is thought that the protein could be

XX CC involved in neurological disorders, infertility, spontaneous abortion,

XX CC neonatal chromosome disorders, aneuploidy and cancers. This is due to its

XX CC function in the movement of microtubules. The protein shows homology to

XX CC the murine KIF1A and KIF1B proteins. The sequences disclosed in the

XX CC invention can be used in the isolation of similar human proteins and in

XX CC vector production. In addition, the biallelic markers shown can be used

XX CC in disease diagnosis and population studies.

XX SQ Sequence 121162 BP; 33272 A; 24108 C; 25842 G; 37919 T; 21 other;

Query Match 4.7%; Score 97.2; DB 21; Length 121162;

Best Local Similarity 74.1%; Pred. No. 3.6e-12;

Matches 123; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

OY 258 AATAGTATGACTCTGTTGCTTTCAGCAGGACATATCTAAATAGGAGCTATACAAG 317

DB 47492 AATAACATAGCAGCGTGCTCGTTCGGCAGCATATCTAAATTTGGAACGATACAGAG 47551

OY 318 AAGATTAGCATGGACTCTGTGCAAGAAATGACACACAAATTTGGAACATTTCCATATATT 377

DB 47552 AAGATTAGCATGGCCCTCGGCAAGATGACGCAAAATTCGGAAGCTTCCATATTTA 47611

KW anti allergic; anti asthmatic; anti inflammatory; olfactory;
KW respiratory active; ds.
XX Homo sapiens.
OS WO20015448-A1.
PN 02-AUG-2001.
XX 17-JAN-2001; 2001WO-US01333.
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184564.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225368.
PR 14-AUG-2000; 2000US-0225370.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231124.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 01-NOV-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.

```
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX Rosen CA, Barash SC, Ruben SM;
PI
XX WPI; 2001-476224/51.
XX
XX Isolated polypeptide for treating, preventing and/ or prognosing
PT disorders related to the respiratory system including respiratory
PT cancers and also for testing and detection e.g. diagnosis -
XX
XX Disclosure; SEQ ID No 1029; 545pp; English.
XX
XX The present invention relates to the isolation of novel human
CC respiratory antigens (AAU17685-AAU17975), and cDNA and genomic
CC sequences encoding for these polypeptides. The sequences of the
CC invention are useful for preventing, treating and/or prognosing
CC disorders related to the respiratory system including throat
CC disorders (e.g. vocal cord paralysis, tonsillitis, and laryngitis),
CC lung disorders e.g. pneumonia, allergic disorders e.g. asthma,
CC pleurisy, cystic fibrosis, emphysema, nose disorders and cancers of
CC the respiratory tissues e.g. lung cancer. The polynucleotide sequences
CC of the invention are useful in gene therapy and antisense therapy.
CC AAS28161-AAS28764 represent genomic sequences encoding for novel
CC human respiratory antigens.
CC
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
XX Sequence 6040 BP; 1446 A; 1488 C; 1420 G; 1686 T; 0 other;
SQ
Query Match 4.6%; Score 95.4; DB 22; Length 6040;
Best Local Similarity 72.8%; Pred. No. 3.5e-12;
Matches 123; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
QY 1136 ATATATAGTATTGGTGTAGATGGATTTTGGCCATGTTGCCAGGCTAGTATTGAAC 1195
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3417 ATTTTATATTTTGTAGTGGAGATGGGTTTGGCCATATTGGCCAGGCTGTGAAC 3358
QY 1196 CTAAGCTAAGCAATCTCTCTGCTGCTGCCATCCCAAAATTTGGGATTACAGGTGAAC 1255
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3357 CTGACCTCAGGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3298
QY 1256 CACTGACCGGCTGATAGCTGGTTTCACTTACTTACTTACTTACTTACTTACTTACT 1304
DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3297 CACTGACCTGCGGCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCT 3249
RESULT 13
AAK84643/c
ID AAK84643 standard; DNA; 4181 BP.
XX
XX AAK84643;
AC
XX
XX 07-NOV-2001 (first entry)
XX
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39455.
DE
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW cytostatic; gene therapy; vaccine; metastasis; ds.
XX
XX Homo sapiens.
XX
XX WO200157182-A2.
XX
XX 09-AUG-2001.
XX
XX 17-JAN-2001; 2001WO-US01354.
XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
```


PR	02-MAR-2000;	2000US-0186350.	PR	02-OCT-2000;	2000US-0237040.
PR	16-MAR-2000;	2000US-0189874.	PR	13-OCT-2000;	2000US-0239935.
PR	17-MAR-2000;	2000US-0190076.	PR	13-OCT-2000;	2000US-0239937.
PR	18-MAR-2000;	2000US-0198123.	PR	20-OCT-2000;	2000US-0240960.
PR	19-MAY-2000;	2000US-0205515.	PR	20-OCT-2000;	2000US-0241221.
PR	07-JUN-2000;	2000US-0209457.	PR	20-OCT-2000;	2000US-0241785.
PR	28-JUN-2000;	2000US-0214886.	PR	20-OCT-2000;	2000US-0241786.
PR	30-JUN-2000;	2000US-0215135.	PR	20-OCT-2000;	2000US-0241787.
PR	07-JUL-2000;	2000US-0216647.	PR	20-OCT-2000;	2000US-0241808.
PR	11-JUL-2000;	2000US-0217487.	PR	20-OCT-2000;	2000US-0241809.
PR	11-JUL-2000;	2000US-0217496.	PR	01-NOV-2000;	2000US-0241826.
PR	14-JUL-2000;	2000US-0218290.	PR	01-NOV-2000;	2000US-0246117.
PR	26-JUL-2000;	2000US-0220963.	PR	08-NOV-2000;	2000US-0246474.
PR	26-JUL-2000;	2000US-0220964.	PR	08-NOV-2000;	2000US-0246475.
PR	14-AUG-2000;	2000US-0224518.	PR	08-NOV-2000;	2000US-0246476.
PR	14-AUG-2000;	2000US-0224519.	PR	08-NOV-2000;	2000US-0246477.
PR	14-AUG-2000;	2000US-0225213.	PR	08-NOV-2000;	2000US-0246478.
PR	14-AUG-2000;	2000US-0225214.	PR	08-NOV-2000;	2000US-0246523.
PR	14-AUG-2000;	2000US-0225266.	PR	08-NOV-2000;	2000US-0246524.
PR	14-AUG-2000;	2000US-0225267.	PR	08-NOV-2000;	2000US-0246525.
PR	14-AUG-2000;	2000US-0225268.	PR	08-NOV-2000;	2000US-0246526.
PR	14-AUG-2000;	2000US-0225270.	PR	08-NOV-2000;	2000US-0246527.
PR	14-AUG-2000;	2000US-0225447.	PR	08-NOV-2000;	2000US-0246528.
PR	14-AUG-2000;	2000US-0225757.	PR	08-NOV-2000;	2000US-0246532.
PR	14-AUG-2000;	2000US-0225758.	PR	08-NOV-2000;	2000US-0246609.
PR	14-AUG-2000;	2000US-0225759.	PR	08-NOV-2000;	2000US-0246610.
PR	18-AUG-2000;	2000US-0226279.	PR	08-NOV-2000;	2000US-0246611.
PR	22-AUG-2000;	2000US-0226681.	PR	08-NOV-2000;	2000US-0246613.
PR	22-AUG-2000;	2000US-0226688.	PR	17-NOV-2000;	2000US-0249207.
PR	22-AUG-2000;	2000US-0227182.	PR	17-NOV-2000;	2000US-0249208.
PR	23-AUG-2000;	2000US-0227009.	PR	17-NOV-2000;	2000US-0249209.
PR	30-AUG-2000;	2000US-0228924.	PR	17-NOV-2000;	2000US-0249210.
PR	01-SEP-2000;	2000US-0229287.	PR	17-NOV-2000;	2000US-0249211.
PR	01-SEP-2000;	2000US-0229343.	PR	17-NOV-2000;	2000US-0249212.
PR	01-SEP-2000;	2000US-0229344.	PR	17-NOV-2000;	2000US-0249213.
PR	01-SEP-2000;	2000US-0229345.	PR	17-NOV-2000;	2000US-0249214.
PR	05-SEP-2000;	2000US-0229509.	PR	17-NOV-2000;	2000US-0249215.
PR	05-SEP-2000;	2000US-0229513.	PR	17-NOV-2000;	2000US-0249216.
PR	06-SEP-2000;	2000US-0230437.	PR	17-NOV-2000;	2000US-0249217.
PR	06-SEP-2000;	2000US-0230438.	PR	17-NOV-2000;	2000US-0249218.
PR	08-SEP-2000;	2000US-0231242.	PR	17-NOV-2000;	2000US-0249244.
PR	08-SEP-2000;	2000US-0231243.	PR	17-NOV-2000;	2000US-0249245.
PR	08-SEP-2000;	2000US-0231244.	PR	17-NOV-2000;	2000US-0249264.
PR	08-SEP-2000;	2000US-0231413.	PR	17-NOV-2000;	2000US-0249265.
PR	08-SEP-2000;	2000US-0231414.	PR	17-NOV-2000;	2000US-0249297.
PR	08-SEP-2000;	2000US-0232080.	PR	17-NOV-2000;	2000US-0249299.
PR	08-SEP-2000;	2000US-0232081.	PR	17-NOV-2000;	2000US-0249300.
PR	12-SEP-2000;	2000US-0231968.	PR	01-DEC-2000;	2000US-0250160.
PR	14-SEP-2000;</				

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 23:31:50 ; Search time 120.734 Seconds
(without alignments)
7582.209 Million cell updates/sec

Title: US-10-005-337A-2
Perfect score: 2074
Sequence: 1 ctgcagcaagtacttaatg.....acaagactccttcagccaac 2074

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2_6/prodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/prodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/prodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/prodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/prodata/2/ina/PCUS_COMB.seq: *
6: /cgn2_6/prodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	92.2	4.4	66804	US-09-740-041-3	Sequence 3, Appli
2	89.4	4.3	6669	US-09-212-971-5	Sequence 5, Appli
3	89.4	4.3	6669	US-08-800-929A-5	Sequence 5, Appli
4	89.4	4.3	6669	US-09-617-053A-5	Sequence 5, Appli
5	89.4	4.3	75395	US-09-984-890-3	Sequence 3, Appli
6	87.8	4.2	59065	US-09-813-817-3	Sequence 3, Appli
7	87.8	4.2	59065	US-09-978-197-3	Sequence 3, Appli
8	86.8	4.2	956	US-09-641-638-36	Sequence 36, Appli
9	86.8	4.2	20674	US-09-641-638-651	Sequence 651, App
10	86.2	4.2	99500	US-09-798-096-10	Sequence 10, Appli
C 11	86	4.1	29629	US-09-729-995-3	Sequence 3, Appli
C 12	85.8	4.1	72604	US-09-268-992-7	Sequence 7, Appli
C 13	85.8	4.1	72604	US-09-657-474-7	Sequence 7, Appli
C 14	85.8	4.1	75395	US-09-984-890-3	Sequence 3, Appli
C 15	85.2	4.1	90541	US-09-759-359A-3	Sequence 3, Appli
C 16	85	4.1	330	US-09-078-294-24	Sequence 24, Appli
C 17	84.6	4.1	4550	US-09-103-663-35	Sequence 35, Appli
18	84.4	4.1	72604	US-09-268-992-7	Sequence 7, Appli
19	84.4	4.1	72604	US-09-657-474-7	Sequence 7, Appli
20	84.2	4.1	1001	US-09-641-638-199	Sequence 199, App
21	84.2	4.1	1001	US-09-641-638-200	Sequence 200, App
C 22	84.2	4.1	3001	US-09-539-333D-142	Sequence 142, App
C 23	84.2	4.1	35060	US-08-814-095-7	Sequence 7, Appli
C 24	84.2	4.1	46718	US-09-816-093-3	Sequence 3, Appli
C 25	84	4.1	53332	US-09-801-861-3	Sequence 3, Appli
C 26	84	4.1	246240	US-08-724-394A-20	Sequence 20, Appli
C 27	84	4.1	246240	US-08-724-394A-21	Sequence 21, Appli

ALIGNMENTS

RESULT 1

US-09-740-041-3
; Sequence 3, Application US/09740041
; Patent No.: 6562593
; GENERAL INFORMATION:
; APPLICANT: MERKULOV, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; FILE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001001
; CURRENT APPLICATION NUMBER: US/09/740,041
; CURRENT FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 66804
; TYPE: DNA
; ORGANISM: Human
US-09-740-041-3

Query Match 4.4%; Score 92.2; DB 4; Length 66804;
Best Local Similarity 73.3%; Pred. No. 3.4e-13;
Matches 118; Conservative 0; Mismatches 43; Indels 0; Gaps 0;

QY	1135	TATATATAGTATTGTGTAGAGATGGGATTTTGGCATTTGCCATTTGCCAGGCTAGTATTGAAC	1194
DB	40127	TAATTTTGTATTGTGTAGAGATAGGATTTTGCATTTGCCATTTGCCAGGCTGGTCTCGAAC	40186
QY	1195	TCCTAAGCTAAGCAATCTTCTGCTCTGCTCCCTCCCAAAATGTTGGATTACAGGTCTAAG	1254
DB	40187	TCCTGACTTAGGTGATCCGCCACCTCGGCCCTCCCAAACTGTGGGATTACAAGCATAAG	40246
QY	1255	CCACTGCACCCGGCTAGCTGGTTTCTATTACTCTATTT	1295
DB	40247	CCACTGTGCTGGCCTTTTTCAGTCTGCTCTGT	40287

RESULT 2

US-09-212-971-5
; Sequence 5, Application US/09212971B
; Patent No.: 6107041
; GENERAL INFORMATION:
; APPLICANT: Korneluk, Robert G
; APPLICANT: Mackenzie, Alexander E
; APPLICANT: Liston, Peter
; APPLICANT: Baird, Stephen
; APPLICANT: Tsang, Benjamin K
; APPLICANT: Pratt, Christine
; TITLE OF INVENTION: DETECTION AND MODULATION OF IAPS AND
; TITLE OF INVENTION: NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE

```
; TITLE OF INVENTION: DISEASE
; FILE REFERENCE: 07891/009002
; CURRENT APPLICATION NUMBER: US/09/212,971B
; CURRENT FILING DATE: 1998-12-16
; EARLIER APPLICATION NUMBER: 60/017,354
; EARLIER FILING DATE: 1996-04-26
; EARLIER APPLICATION NUMBER: 60/030,590
; EARLIER FILING DATE: 1996-11-14
; EARLIER APPLICATION NUMBER: 08/800,929
; EARLIER FILING DATE: 1997-02-13
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 6669
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variation
; LOCATION: (3677)...(3951)
; OTHER INFORMATION: n can be any nucleotide
US-09-212-971-5

Query Match      4.3%; Score 89.4; DB 3; Length 6669;
Best Local Similarity 73.1%; Pred. No. 7.2e-13;
Matches 128; Conservative 0; Mismatches 46; Indels 1; Gaps 1;

QY 1135 TATATATAGTATTGTGGTAGAGATGGGATTTGCCCATGTTGCCAGGCTAGTATTGAAC 1194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1230 TAATTTTGTATTTTAGTAGAGATGGGTTTCACCATTTTGGCCAGGCTGGTTTGAAC 1289
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1195 TCCTAAGCT-AAGCAATCTTCCTGCTCTGCCCTCCCAAAATGTTGGATTACAGGTGTAA 1253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1290 TCGTGACCTCAAGTAGATCCACTGTCTGTGGCTCCCAAAATGTTGGATTACAGGTGTGA 1349
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1254 GCCACTGCACCGGCTGATAGCTGGTTTCATTTACTTCTATTTCTTGACCACCTCTG 1308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1350 GCCACTGCACCGGCTGATAGCTGGTTTCATTTACTTCTATTTCTTGACCACCTCTG 1404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
US-08-800-929A-5
; Sequence 5, Application US/08800929A
; Patent No. 6133437
; GENERAL INFORMATION:
; APPLICANT: Korneluk, Robert G
; APPLICANT: Mackenzie, Alexander E
; APPLICANT: Liston, Peter
; APPLICANT: Baird, Stephen
; APPLICANT: Tsang, Benjamin K
; APPLICANT: Pratt, Christine
; TITLE OF INVENTION: DETECTION AND MODULATION OF
; TITLE OF INVENTION: IAPS AND NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Clark & Elbing LLP
; STREET: 176 Federal Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/800,929A
; FILING DATE: 13-FEB-1997
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/030,590
; FILING DATE: 14-NOV-1996
```

```
; APPLICATION NUMBER: 60/017,354
; FILING DATE: 26-APR-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Bleker-Brady, Kristina
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: 07891/009001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-428-0200
; TELEFAX: 617-428-7045
; TELEX:
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6669 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-800-929A-5

Query Match      4.3%; Score 89.4; DB 3; Length 6669;
Best Local Similarity 73.1%; Pred. No. 7.2e-13;
Matches 128; Conservative 0; Mismatches 46; Indels 1; Gaps 1;

QY 1135 TATATATAGTATTGTGGTAGAGATGGGATTTGCCCATGTTGCCAGGCTAGTATTGAAC 1194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1230 TAATTTTGTATTTTAGTAGAGATGGGTTTCACCATTTTGGCCAGGCTGGTTTGAAC 1289
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1195 TCCTAAGCT-AAGCAATCTTCCTGCTCTGCCCTCCCAAAATGTTGGATTACAGGTGTAA 1253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1290 TCGTGACCTCAAGTAGATCCACTGTCTGTGGCTCCCAAAATGTTGGATTACAGGTGTGA 1349
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 1254 GCCACTGCACCGGCTGATAGCTGGTTTCATTTACTTCTATTTCTTGACCACCTCTG 1308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1350 GCCACTGCACCGGCTGATAGCTGGTTTCATTTACTTCTATTTCTTGACCACCTCTG 1404
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
US-09-617-053A-5
; Sequence 5, Application US/09617053A
; Patent No. 6300492
; GENERAL INFORMATION:
; APPLICANT: Korneluk, Robert G
; APPLICANT: Mackenzie, Alexander E
; APPLICANT: Liston, Peter
; APPLICANT: Baird, Stephen
; APPLICANT: Tsang, Benjamin K
; APPLICANT: Pratt, Christine
; TITLE OF INVENTION: DETECTION AND MODULATION OF IAPS AND
; TITLE OF INVENTION: NAIP FOR THE DIAGNOSIS AND TREATMENT OF PROLIFERATIVE
; TITLE OF INVENTION: DISEASE
; FILE REFERENCE: 07891/009003
; CURRENT APPLICATION NUMBER: US/09/617,053A
; CURRENT FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: US 08/800,929
; PRIOR FILING DATE: 1997-02-13
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 6669
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: variation
; LOCATION: (3677)...(3951)
; OTHER INFORMATION: n can be any nucleotide
US-09-617-053A-5

Query Match      4.3%; Score 89.4; DB 4; Length 6669;
Best Local Similarity 73.1%; Pred. No. 7.2e-13;
Matches 128; Conservative 0; Mismatches 46; Indels 1; Gaps 1;

QY 1135 TATATATAGTATTGTGGTAGAGATGGGATTTGCCCATGTTGCCAGGCTAGTATTGAAC 1194
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Db 1230 TAATTTTGTATTTTGTAGATAGAGTGGGTTTCACCAATTTTGGCCAGGCTGGTTTGAAC 1289
QY 1195 TCCTAAGCT-AAGCAATCTTCTGTCTGCTCCCAAAATGTTGGGATTACAGGTGTAA 1253
Db 1290 TCCTGACCTCAAGTATCCACTGTCTTGCCCTCCCAAAATGCTGGATTACAGGCGTGA 1349
QY 1254 GCCACTGCACCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTGACCACCTCTG 1308
Db 1350 GCCACTGCACCGGCTGATAGGCTCTGTTTTTATCTCTTTTTTGGCCCTCTACAGTG 1404

RESULT 5

US-09-984-890-3
; Sequence 3, Application US/09984890
; Patent No. 6492156
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL001306
; CURRENT APPLICATION NUMBER: US/09/984,890
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 75395
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(75395)
; OTHER INFORMATION: n = A,T,C or G
US-09-984-890-3

Query Match 4.3%; Score 89.4; DB 4; Length 75395;
Best Local Similarity 72.1%; Pred. No. 1.8e-12;
Matches 129; Conservative 1; Mismatches 48; Indels 1; Gaps 1;
QY 1090 CAGAGTCCTGCCATTCCTAATATCTGCTCATTTTBTCAATATATATATAGTATTG 1149
Db 13177 CAGGCACCTGCCACCATCTGCTCAATTTTTTTTTTTTTTTTTTTTTTTTTTTT 13236
QY 1150 TGGTAGAGTGGGATTTGCCATGTTGCCAGGCTAGTATTCACCTCCTAAGCTAAGCAA 1209
Db 13237 TAGTAGAGATGGGTTTCACCATGTTGACCAAGCTGCTTGAACCTCCTGACCTCA-TGA 13295
QY 1210 TCTTCTGCTCTGCTCCCAAAATGTTGGGATTACAGGTGTAAGCCACTGCACCGGC 1268
Db 13296 TCTGCCACCTCGGCTCCCAAGTCTGGGATTACAGGCGTGAGCCACTGCACCGGCC 13354

RESULT 6

US-09-813-817-3
; Sequence 3, Application US/09813817
; Patent No. 6340583
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL001178
; CURRENT APPLICATION NUMBER: US/09/813,817
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 59065
; TYPE: DNA
; ORGANISM: Human
US-09-813-817-3

Query Match 4.2%; Score 87.8; DB 4; Length 59065;

Best Local Similarity 78.1%; Pred. No. 4.2e-12;
Matches 118; Conservative 0; Mismatches 32; Indels 1; Gaps 1;
QY 1135 TATATATAGTATTGTTGGTAGAGATGGGATTTTGGCATGTTGCCAGGCTAGTATTGAAC 1194
Db 11804 TAATTTTATATTTTGTAGAGATGGGGTTTGGCATGTTGCCAAGCTAGTCTCGAAC 11863
QY 1195 TCCTAAGCT-AAGCAATCTTCTGTCTCTGCTCCCAAAATGTTGGGATTACAGGTGTAA 1253
Db 11864 TCCTGACCTCAAGTATCCACCTGCTCGGCTCCCAAGTCTGGGATTACAGGTGTGA 11923
QY 1254 GCCACTGCACCGGCTGATAGCTGGTTTCAT 1284
Db 11924 GCCACTGCACCTGGCTGGAAGGAGTGATCTT 11954

RESULT 7

US-09-978-197-3
; Sequence 3, Application US/09978197
; Patent No. 6403353
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL001178DIV
; CURRENT APPLICATION NUMBER: US/09/978,197
; PRIOR FILING DATE: 2001-10-17
; PRIOR FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 59065
; TYPE: DNA
; ORGANISM: Human
US-09-978-197-3

Query Match 4.2%; Score 87.8; DB 4; Length 59065;
Best Local Similarity 78.1%; Pred. No. 4.2e-12;
Matches 118; Conservative 0; Mismatches 32; Indels 1; Gaps 1;
QY 1135 TATATATAGTATTGTTGGTAGAGATGGGATTTTGGCATGTTGCCAGGCTAGTATTGAAC 1194
Db 11804 TAATTTTATATTTTGTAGAGATGGGGTTTGGCATGTTGCCAAGCTAGTCTCGAAC 11863
QY 1195 TCCTAAGCT-AAGCAATCTTCTGTCTCTGCTCCCAAAATGTTGGGATTACAGGTGTAA 1253
Db 11864 TCCTGACCTCAAGTATCCACCTGCTCGGCTCCCAAGTCTGGGATTACAGGTGTGA 11923
QY 1254 GCCACTGCACCGGCTGATAGCTGGTTTCAT 1284
Db 11924 GCCACTGCACCTGGCTGGAAGGAGTGATCTT 11954

RESULT 8

US-09-641-638-36
; Sequence 36, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bougueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; FILE REFERENCE: GENSET.051CPI
; CURRENT APPLICATION NUMBER: US/09/641,638
; PRIOR FILING DATE: 2000-08-16
; PRIOR FILING DATE: 2000-02-11
; PRIOR FILING DATE: 2000-02-11
; PRIOR FILING DATE: 1999-05-07


```
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; PRIOR FILING DATE: 1999-02-12
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 36
; LENGTH: 956
; TYPE: DNA
; ORGANISM: Homo Sapiens
; FEATURE:
; LOCATION: 478
; OTHER INFORMATION: 10-349-47 : polymorphic base C or T
; NAME/KEY: misc_binding
; LOCATION: 459..477
; OTHER INFORMATION: 10-349-47.mis1, potential
; NAME/KEY: misc_binding
; LOCATION: 479..498
; OTHER INFORMATION: 10-349-47.mis2, potential complement
; NAME/KEY: primer_bind
; LOCATION: 432..451
; OTHER INFORMATION: upstream amplification primer
; NAME/KEY: primer_bind
; LOCATION: 829..848
; OTHER INFORMATION: downstream amplification primer, complement
; NAME/KEY: misc_binding
; LOCATION: 466..490
; OTHER INFORMATION: 10-349-47 potential probe
; US-09-641-638-36

Query Match
Best Local Similarity 4.2%; Score 86.8; DB 4; Length 956;
Matches 118; Conservative 1; Mismatches 34; Indels 1; Gaps 1;

QY 1116 GTCCTATTTTBTTCATATATATATAGTATTTGTGTAGAGATGGGATTTTGCCTATGT 1175
Db 9 GCACCATGTGGCGCTGGCTAATTTTGTATTTTGTAGATGGGTTTGCCTATAT 68

QY 1176 GCCAGGCTAGTATTAAGTCTTAAGCT-AAGCAATCTTCTGTCTCTCCCAAAAT 1234
Db 69 GCCAGGCTGTCTTGAAGCTTCTGGCTCAAGTAATCTGCTCGCTCGGCTCCCAAGT 128

QY 1235 GTTGGATTACAGGTCTAGCCACTGCACCCGC 1268
Db 129 GCTGGGATTACAGCTGTGAGCCACTGCACCTGC 162

RESULT 9
US-09-641-638-651
; Sequence 651, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
; APPLICANT: Blumenfeld, Marta
; APPLICANT: Bouqueleret, Lydie
; APPLICANT: Chumakov, Ilya
; APPLICANT: Cohen, Annick
; TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
; FILE REFERENCE: GENSET.051CPI
; CURRENT APPLICATION NUMBER: US/09/641,638
; CURRENT FILING DATE: 2000-08-16
; PRIOR APPLICATION NUMBER: US 09/502,330
; PRIOR FILING DATE: 2000-02-11
; PRIOR APPLICATION NUMBER: US 60/133,200
; PRIOR FILING DATE: 1999-05-07
; PRIOR APPLICATION NUMBER: US 09/275,267
; PRIOR FILING DATE: 1999-03-23
; PRIOR APPLICATION NUMBER: US 60/119,917
; NUMBER OF SEQ ID NOS: 1304
; SOFTWARE: Patent.pm
; SEQ ID NO 651
```


Db 32693 ATTTTGTGTTTATTAGAGTGGGTTTCCACATGTTGGCCAGGCTGCTCGAACT 32752
QY 1196 CTTAAGCT-AAGCAATCTTCTGTCTCTGCTCCCAAAATGTTGGGATTACAGGTGAAG 1254
Db 32753 CCTGACCTCAAGCATTCACCTGCTGGGCTCCCAAGTGTGGGATTACAGGCATGAG 32812
QY 1255 CCACTGACCCGGCTGATAGCTGTTTCATTTACTCTAT 1293
Db 32813 CCACTGACCCGGCTCTTCCCTTTTCAATTATGCTGT 32851

RESULT 11

US-09-729-995-3/c
; Sequence 3, Application US/09729995
; Patent No. 6426206
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CLO00904
; CURRENT APPLICATION NUMBER: US/09/729,995
; CURRENT FILING DATE: 2000-12-06
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 29629
; TYPE: DNA
; ORGANISM: Human
US-09-729-995-3

Query Match 4.1%; Score 86; DB 4; Length 29629;
Best Local Similarity 77.5%; Pred. No. 9.2e-12;
Matches 104; Conservative 0; Mismatches 30; Indels 0; Gaps 0;

QY 1135 TATATAGTATTGTTGGTAGAGATGGGATTTTGGCCATGTTGCCAGGCTAGTATTGAAC 1194
Db 16878 TAATTTTGTATTTTATAGATAGATGGGTTTCCACATGTTGCCAGGCTGTTCTTAAC 16819
QY 1195 TCCTAAGCTAAGCAATCTTCTGTCTCTGCTCCCAAAATGTTGGGATTACAGGTGAAG 1254
Db 16818 TCCTAAGCTAAGTATCTGCTGCCACCTCAGCTCCCAAGTGTGGGATTACAGGCATGAG 16759
QY 1255 CCACTGACCCGGC 1268
Db 16758 CCACGACCCGGC 16745

RESULT 12

US-09-268-992-7/c
; Sequence 7, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 72604

; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified base
; LOCATION: all n positions
; OTHER INFORMATION: n=a, c, g, or t
US-09-268-992-7

Query Match 4.1%; Score 85.8; DB 4; Length 72604;
Best Local Similarity 72.5%; Pred. No. 1.5e-11;
Matches 111; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 1135 TATATAGTATTGTTGGTAGAGATGGGATTTTGGCCATGTTGCCAGGCTAGTATTGAAC 1194
Db 57580 TTTTGTGTTTATTTTAAATAGATAGGTTTCCACATGTTGGTCAAGGCTGGTCTCGAAC 57521
QY 1195 TCCTAAGCTAAGCAATCTTCTGTCTCTGCTCCCAAAATGTTGGGATTACAGGTGAAG 1254
Db 57520 TCCTGACCTCGTCGATCCGCTTGGCTCCCAAAATGCTGGGATTACAGGCATGAG 57461
QY 1255 CCACTGACCCGGCTGATAGCTGGTTCATTTA 1287
Db 57460 CCACGACCCGGCTGATGAAGCATGTTTAA 57428

RESULT 13

US-09-657-474-7/c
; Sequence 7, Application US/09657474
; Patent No. 6399762
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/657,474
; CURRENT FILING DATE: 2000-09-07
; PRIOR APPLICATION NUMBER: 09/268,992
; PRIOR FILING DATE: 1999-03-16
; PRIOR APPLICATION NUMBER: 09/236,134
; PRIOR FILING DATE: 1999-01-22
; PRIOR APPLICATION NUMBER: 60/106,056
; PRIOR FILING DATE: 1998-10-28
; PRIOR APPLICATION NUMBER: 60/088,312
; PRIOR FILING DATE: 1998-06-05
; PRIOR APPLICATION NUMBER: 60/078,044
; PRIOR FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 72604
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified base
; LOCATION: all n positions
; OTHER INFORMATION: n=a, c, g, or t
US-09-657-474-7

Query Match 4.1%; Score 85.8; DB 4; Length 72604;
Best Local Similarity 72.5%; Pred. No. 1.5e-11;
Matches 111; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 1135 TATATAGTATTGTTGGTAGAGATGGGATTTTGGCCATGTTGCCAGGCTAGTATTGAAC 1194
Db 57580 TTTTGTGTTTATTTTAAATAGATAGGTTTCCACATGTTGGTCAAGGCTGGTCTCGAAC 57521
QY 1195 TCCTAAGCTAAGCAATCTTCTGTCTCTGCTCCCAAAATGTTGGGATTACAGGTGAAG 1254
Db 57520 TCCTGACCTCGTCGATCCGCTTGGCTCCCAAAATGCTGGGATTACAGGCATGAG 57461
QY 1255 CCACTGACCCGGCTGATAGCTGGTTCATTTA 1287

Db 57460 CCACCACCCGGCTGTATGAAGCATGTTTTTA 57428

RESULT 14
US-09-984-890-3/c
; Sequence 3, Application US/09984890
; Patent No. 6492156
; GENERAL INFORMATION:
; APPLICANT: YAN, Chunhua et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001306
; CURRENT APPLICATION NUMBER: US/09/984,890
; CURRENT FILING DATE: 2001-10-31
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 75395
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (1)...(75395)
; OTHER INFORMATION: n = A,T,C or G
US-09-984-890-3

Query Match 4.1%; Score 85.8; DB 4; Length 75395;
Best Local Similarity 78.6%; Pred. No. 1.5e-11;
Matches 114; Conservative 1; Mismatches 29; Indels 1; Gaps 1;
QY 1125 TTBTCAATATATATAGTATTGTTGGTAGAGATGGGATTTGCCCATGTTGCCCAGGCT 1184
Db 44960 TAGTTTTTATTATTTCATTTTTTTTGTAGAGATGGCATCTTGGCCATCTTGGCCAGGCT 44901
QY 1185 AGTATTGAACCTCTAAGCT-AAGCAATCTTCTGCTCTGCCCTCCCAAAATGTTGGGATT 1243
Db 44900 GGCCCTCAAACTCTAGGCTCAAGCAATCTTCTGCTCTGCCCTCCCAAAAGTGTGGGATT 44841
QY 1244 ACAGGTGTAAGCCACTGCCCCGGC 1268
Db 44840 ACAGGCTGAGCCACCACCCAGC 44816

RESULT 15
US-09-759-359A-3/c
; Sequence 3, Application US/09759359A
; Patent No. 6492153
; GENERAL INFORMATION:
; APPLICANT: ABU-THREIDEH, Jane et al
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; TITLE OF INVENTION: ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; TITLE OF INVENTION: THEREOF
; FILE REFERENCE: CL001043
; CURRENT APPLICATION NUMBER: US/09/759,359A
; CURRENT FILING DATE: 2001-01-16
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 90541
; TYPE: DNA
; ORGANISM: Human
US-09-759-359A-3

Query Match 4.1%; Score 85.2; DB 4; Length 90541;
Best Local Similarity 76.0%; Pred. No. 2.2e-11;
Matches 117; Conservative 1; Mismatches 35; Indels 1; Gaps 1;
QY 1116 GTCTCATTTTBTTCATATATATATAGTATTGTTGGTAGAGATGGGATTTGCCCATGTT 1175
Db 23174 GGTAAATTTTGTGTTTTTTATTATTATTATTTAAAGTAGAGATGGGATTTGCCCATGTT 23115
QY 1176 GCCCAGGCTAGTATTGAACCTCTAAGCT-AAGCAATCTTCTGCTCTGCCCTCCCAAAAT 1234

Db 23114 GGCCAGGCTGGTCTCAAACTCTTAGCCTCAAGCAATCCACCTGCTTGCCCTCCCAAGT 23055
QY 1235 GTTGGGATTACAGGTGTGAAGCCCACTGCACCCGGC 1268
Db 23054 GCTGGATTACAGGCGTGAGCCACACAGTGCCTGCG 23021

Search completed: November 18, 2003, 07:06:35
Job time : 139.734 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 23:33:10 ; Search time 992.544 Seconds
(without alignments)
6829.176 Million cell updates/sec

Title: US-10-005-337A-2

Perfect score: 2074

Sequence: 1 cgcgcagcagttacttaag.....acaagactcttcagcgaac 2074

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 4339922

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA.*

- 1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/prodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/prodata/1/pubpna/US09A_PUBCOMB.seq.*
- 10: /cgn2_6/prodata/1/pubpna/US09B_PUBCOMB.seq.*
- 11: /cgn2_6/prodata/1/pubpna/US09C_PUBCOMB.seq.*
- 12: /cgn2_6/prodata/1/pubpna/US09_NEW_PUB.seq.*
- 13: /cgn2_6/prodata/1/pubpna/US10A_PUBCOMB.seq.*
- 14: /cgn2_6/prodata/1/pubpna/US10B_PUBCOMB.seq.*
- 15: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq.*
- 16: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq.*
- 17: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2071.6	99.9	2074	14	US-10-005-337A-2
2	367.8	17.7	2358	14	US-10-005-337A-1
3	144.4	7.0	1889	12	US-09-974-298-182
4	144.4	7.0	1889	12	US-10-240-965-158
5	144.4	7.0	1889	12	US-10-252-157-421
6	141.4	6.8	1325	14	US-10-084-817-217
7	129.4	6.2	184	9	US-09-923-876-5218
8	95.4	4.6	6040	9	US-09-764-860-1029
9	95.4	4.6	6040	14	US-10-074-095-1029
10	94.8	4.6	300000	14	US-10-262-552-33
11	93.8	4.5	119040	11	US-09-911-077A-19
12	93.2	4.5	32220	10	US-09-764-877-3933
13	92.6	4.5	555	12	US-10-027-632-95562
14	92.6	4.5	555	12	US-10-027-632-95563
15	92.6	4.5	555	13	US-10-027-632-95562
16	92.6	4.5	555	13	US-10-027-632-95563

17	92.2	4.4	66804	9	US-09-740-041-3	Sequence 3, Appli
18	92.2	4.4	66804	12	US-10-389-967-3	Sequence 3, Appli
19	91	4.4	461	11	US-09-918-995-35181	Sequence 35181, A
20	91	4.4	763	12	US-10-027-632-27383	Sequence 27383, A
21	91	4.4	763	12	US-10-027-632-27384	Sequence 27384, A
22	91	4.4	763	13	US-10-027-632-27383	Sequence 27383, A
23	91	4.4	763	13	US-10-027-632-27384	Sequence 27384, A
24	91	4.4	56737	10	US-09-782-378A-17	Sequence 17, Appli
25	90.4	4.3	19820	10	US-09-764-877-2713	Sequence 2713, Ap
26	90.2	4.3	701	12	US-10-027-632-113259	Sequence 113259,
27	90.2	4.3	701	12	US-10-027-632-113260	Sequence 113260,
28	90.2	4.3	701	13	US-10-027-632-113259	Sequence 113259,
29	90.2	4.3	701	13	US-10-027-632-113260	Sequence 113260,
30	90.2	4.3	705	12	US-10-027-632-268052	Sequence 268052,
31	90.2	4.3	705	12	US-10-027-632-268053	Sequence 268053,
32	90.2	4.3	705	13	US-10-027-632-268052	Sequence 268052,
33	90.2	4.3	705	13	US-10-027-632-268053	Sequence 268053,
34	90	4.3	625	12	US-10-027-632-102687	Sequence 102687,
35	90	4.3	625	12	US-10-027-632-102688	Sequence 102688,
36	90	4.3	625	12	US-10-027-632-102689	Sequence 102689,
37	90	4.3	625	13	US-10-027-632-102687	Sequence 102687,
38	90	4.3	625	13	US-10-027-632-102688	Sequence 102688,
39	90	4.3	625	13	US-10-027-632-102689	Sequence 102689,
40	90	4.3	715	12	US-10-027-632-110575	Sequence 110575,
41	90	4.3	715	12	US-10-027-632-110576	Sequence 110576,
42	90	4.3	715	13	US-10-027-632-110575	Sequence 110575,
43	90	4.3	715	13	US-10-027-632-110576	Sequence 110576,
44	89.8	4.3	86592	12	US-10-211-160-1	Sequence 1, Appli
45	89.8	4.3	86592	12	US-10-051-681A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-10-005-337A-2
; Sequence 2, Application US/10005337A
; Publication No. US20030039984A1
; GENERAL INFORMATION:
; APPLICANT: BENOIT, Patrick
; APPLICANT: SCHWARTZ, Bertrand
; APPLICANT: BRANELLEC, Didier
; APPLICANT: CHIEN, Kenneth R.
; TITLE OF INVENTION: SEQUENCES UPSTREAM OF THE CARP GENE, VECTORS CONTAINING
; TITLE OF INVENTION: THEM AND USES THEREOF
; FILE REFERENCE: 03806.0530-00000
; CURRENT APPLICATION NUMBER: US/10/005,337A
; CURRENT FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: US 60/251,582
; PRIOR FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2074
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-005-337A-2

Query Match	99.9%;	Score	2071.6;	DB	14;	Length	2074;	
Best Local Similarity	100.0%;	Pred. No.	0;					
Matches 2074;	Conservative	0;	Mismatches	0;	Indels	0;	Gaps	0;
QY	1	CTGCACGAGTTACTTAATGTTTTTTGCTCAGCATCCTCTCTGTAAATAATGAGAGCATTA	60					
Db	1	CTGCACGAGTTACTTAATGTTTTTTGCTCAGCATCCTCTCTGTAAATAATGAGAGCATTA	60					
QY	61	GTCTTGCTCCAACTTCGAGGGCATGGACAGCTCTGGGATTTTCATATCCAAAGACCCCTTAAA	120					
Db	61	GTCTTGCTCCAACTTCGAGGGCATGGACAGCTCTGGGATTTTCATATCCAAAGACCCCTTAAA	120					
QY	121	CATCCACAGTCTTTCCTCCCAACAACTTCTCTCTCTTAATACCTCCCTCAGTTTGGGTGAG	180					
Db	121	CATCCACAGTCTTTCCTCCCAACAACTTCTCTCTCTTAATACCTCCCTCAGTTTGGGTGAG	180					

181 GCCTGGNACAAAAGGCGATACGAAATGGTGAAGAAAGTGTCCATGACTACTTCTGACTTA 240
181 GCCTGGNACAAAAGGCGATACGAAATGGTGAAGAAAGTGTCCATGACTACTTCTGACTTA 240
241 GATTGAAGAGACCAATGAAATAGTAATGACTCTGTGTTGCTTTCAGCAGGACATATCTAA 300
241 GATTGAAGAGACCAATGAAATAGTAATGACTCTGTGTTGCTTTCAGCAGGACATATCTAA 300
301 ATAGGAGCTATACAAAGAAAGTATAGCATGGACTCTGTGCAAGAAATGACACACAAAATTTGT 360
301 ATAGGAGCTATACAAAGAAAGTATAGCATGGACTCTGTGCAAGAAATGACACACAAAATTTGT 360
361 GAAACATTCATATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 420
361 GAAACATTCATATATATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 420
421 AAAATAGTAGTAGTGTGCTCCATCTCAAGAAAGCCAGGAGATTTCTTTATTTACCC 480
421 AAAATAGTAGTAGTGTGCTCCATCTCAAGAAAGCCAGGAGATTTCTTTATTTACCC 480
481 CCTTTAAGATAGAAATATTAGGAGACCGGAAACATATGATACAGGAGTACTGGAGGGTCC 540
481 CCTTTAAGATAGAAATATTAGGAGACCGGAAACATATGATACAGGAGTACTGGAGGGTCC 540
541 CTCTTTGTCGAATGTTTGTCTTGGGTGGGAGTGCATGCTCTCAAGTTTCAGAAAC 600
541 CTCTTTGTCGAATGTTTGTCTTGGGTGGGAGTGCATGCTCTCAAGTTTCAGAAAC 600
601 ACCATCCACTGACTGAGCATTCAGGGGCAAGAGGAGAAATGGCAGCCACATTTGTTGATT 660
601 ACCATCCACTGACTGAGCATTCAGGGGCAAGAGGAGAAATGGCAGCCACATTTGTTGATT 660
661 GGGTGAAGTTTGGGAGAAATAGACACACAAAGGTCAAACTACTTCTTAATTAACACTT 720
661 GGGTGAAGTTTGGGAGAAATAGACACACAAAGGTCAAACTACTTCTTAATTAACACTT 720
721 CCTTCGATCAAAATCCCTTCTCCGATTTCTCTCTGCTGCTTTTACTTSKARAAACCC 780
721 CCTTCGATCAAAATCCCTTCTCCGATTTCTCTCTGCTGCTTTTACTTSKARAAACCC 780
781 AGTTTTCTGAACTATAAAATACCCGAGTATGTTTACATAAATTTACACCTCAAGA 840
781 AGTTTTCTGAACTATAAAATACCCGAGTATGTTTACATAAATTTACACCTCAAGA 840
841 TTAGAAACCAAGAAATAGAGACCTTTTCAACCCCTTCCGGAAGCAAGTGCATTTCCCTCC 900
841 TTAGAAACCAAGAAATAGAGACCTTTTCAACCCCTTCCGGAAGCAAGTGCATTTCCCTCC 900
901 AGCCAGGTGCTCAAAATCTTTGATGCAATCAGAAATCATCTGGGTGCTTTKAAATTCAGATG 960
901 AGCCAGGTGCTCAAAATCTTTGATGCAATCAGAAATCATCTGGGTGCTTTKAAATTCAGATG 960
961 ATTCTACGAGTTACCAATTAATCAACTCAGAAATCCCTGAGTGGGCGCAGGATCTGTA 1020
961 ATTCTACGAGTTACCAATTAATCAACTCAGAAATCCCTGAGTGGGCGCAGGATCTGTA 1020
1021 TTTCTGACAAAGCTCCACAGGTGATTTCTTTTCCCCACAGCATTTGAGAACTTCAGCTCAA 1080
1021 TTTCTGACAAAGCTCCACAGGTGATTTCTTTTCCCCACAGCATTTGAGAACTTCAGCTCAA 1080
1081 TGACCTAATCAGAGTCTCGCATTTGCTATATCTGCTCTCAATTTTBTTCATATATATATA 1140
1081 TGACCTAATCAGAGTCTCGCATTTGCTATATCTGCTCTCAATTTTBTTCATATATATATA 1140
1141 TAGTATTTGGTGAAGATGGGATTTTGCCATGTTGCCAGGCTAGTATTGAACTCTCTAA 1200
1141 TAGTATTTGGTGAAGATGGGATTTTGCCATGTTGCCAGGCTAGTATTGAACTCTCTAA 1200
1201 GCTAAGCAATCTTCTGCTCTGCTCCCTCCCAAAATGTTGGGATTTACAGGTGTAAAGCCACTG 1260
1201 GCTAAGCAATCTTCTGCTCTGCTCCCTCCCAAAATGTTGGGATTTACAGGTGTAAAGCCACTG 1260

1261 CACCCGGCTGATAGCTGGTTTCATTTTACTCTATTTCTTGACCACTCTGATCCATTTTGAA 1320
1261 CACCCGGCTGATAGCTGGTTTCATTTTACTCTATTTCTTGACCACTCTGATCCATTTTGAA 1320
1321 GTAAAAATGCTCCAATTTATTTATGCTGTTTTAGAACACGGTAAAGCATGTCATGTGCTAAATG 1380
1321 GTAAAAATGCTCCAATTTATTTATGCTGTTTTAGAACACGGTAAAGCATGTCATGTGCTAAATG 1380
1381 GCAGTGAATCATATAAAGAAAGTGCATTAATGATGCTTTCAATGCTTTATATAATGATG 1440
1381 GCAGTGAATCATATAAAGAAAGTGCATTAATGATGCTTTCAATGCTTTATATAATGATG 1440
1441 GTAAAGTGCATGTCATGGGGCTATTAGCCAGACATCACTCCAAAGAAATTTCCAAACA 1500
1441 GTAAAGTGCATGTCATGGGGCTATTAGCCAGACATCACTCCAAAGAAATTTCCAAACA 1500
1501 GATATAGACAAAGTGCCTTTTAGGCCCCAGATCCCTTCCCTCAGGCTGTTTACCAGGGAA 1560
1501 GATATAGACAAAGTGCCTTTTAGGCCCCAGATCCCTTCCCTCAGGCTGTTTACCAGGGAA 1560
1561 TAGGATGTCCTGGGACAAAGTTTCCCTTAAAGTGAAGTGTGATTAAGTCTGCTTATCAGAAA 1620
1561 TAGGATGTCCTGGGACAAAGTTTCCCTTAAAGTGAAGTGTGATTAAGTCTGCTTATCAGAAA 1620
1621 GATATTACTGGGGTGTGATATGATAGGCGATCTACATTTTCTTGATAGGTAGTCATATGA 1680
1621 GATATTACTGGGGTGTGATATGATAGGCGATCTACATTTTCTTGATAGGTAGTCATATGA 1680
1681 AAGCTGACAAAAGAAAAAGGCGAGTGTGTGCAATGTCAACAGACAGCTGTCCCTT 1740
1681 AAGCTGACAAAAGAAAAAGGCGAGTGTGTGCAATGTCAACAGACAGCTGTCCCTT 1740
1741 GACTCTTGACAAATAGGATGACTTGCTTGTGAGCGATGTGATCAACCAAGGAATG 1800
1741 GACTCTTGACAAATAGGATGACTTGCTTGTGAGCGATGTGATCAACCAAGGAATG 1800
1801 GGCCTCTCAATTTCTTCTGATTCACATATTCAGCAGGTTAGCTGCTCCCTCCCTCC 1860
1801 GGCCTCTCAATTTCTTCTGATTCACATATTCAGCAGGTTAGCTGCTCCCTCCCTCC 1860
1861 TCTTCAGCTTCCAGACACTGAGTCTGGAATGAAATTCACCTGCTCTGAGTTGGCTCC 1920
1861 TCTTCAGCTTCCAGACACTGAGTCTGGAATGAAATTCACCTGCTCTGAGTTGGCTCC 1920
1921 TAATGGGGCGGAGTGTACTTGGTTCCAGGTTGGAAGATTAATCAACCGGCGCCCA 1980
1921 TAATGGGGCGGAGTGTACTTGGTTCCAGGTTGGAAGATTAATCAACCGGCGCCCA 1980
1981 GCTATATAAGCTGACCGGTGTGGGGGCGCCAGCAGGCGCACTCCAGGATTCCTTCCA 2040
1981 GCTATATAAGCTGACCGGTGTGGGGGCGCCAGCAGGCGCACTCCAGGATTCCTTCCA 2040
2041 CGACAGAAAAACATACAAGACTCCTTCAGCCCAAC 2074
2041 CGACAGAAAAACATACAAGACTCCTTCAGCCCAAC 2074

RESULT 2

US-10-005-337A-1
; Sequence 1: Application US/10005337A
; Publication No. US2003039984A1
; GENERAL INFORMATION:
; APPLICANT: BENOIT, Patrick
; APPLICANT: SCHWARTZ, Bertrand
; APPLICANT: BRANELLEC, Didier
; APPLICANT: CHIEN, Kenneth R.
; TITLE OF INVENTION: SEQUENCES UPSTREAM OF THE CARP GENE, VECTORS CONTAINING
; FILE OF INVENTION: THEM AND USES THEREOF
; FILE REFERENCE: 03806.0530-00000
; CURRENT APPLICATION NUMBER: US/10/005,337A
; CURRENT FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: US 60/251,582
; PRIOR FILING DATE: 2000-12-07

	Query Match	7.0%;	Score 144.4;	DB 12;	Length 1889;
	Best Local Similarity	95.3%;	Pred. No. 2.6e-26;		
	Matches 145;	Conservative 0;	Mismatches 1;	Indels 0;	Gaps 0;
Qy	1929	GCGGGAGTGTTACTTCGGTTCCTCCAGGTTGGAAGATTATCTCACCGGCCCCAGCTATATA	1988		
Db	5	GAGGGAGTGTTACTTCGGTTCCTCCAGGTTGGAAGATTATCTCACCGGCCCCAGCTATATA	64		
Qy	1989	AGCTTGACCGGTGTGTGGAGGGGCCAGCAGGGCCAACTCAGGGATTCTTCCACGACAGAA	2048		
Db	65	AGCTGACCGGTGTGTGGAGGGGCCAGCAGGGCCAACTCAGGGATTCTTCCACGACAGAA	124		
Qy	2049	AAACATACAAGACTCTCTTCAGCCAAAC	2074		
Db	125	AAACATACAAGACTCTCTTCAGCCAAAC	150		

RESULT 5

```

US-10-252-157-421
; Sequence 421, Application US/10252157
; Publication No. US20030190640A1
; GENERAL INFORMATION:
; APPLICANT: Fatis, Mary
; APPLICANT: Pearson, Cecelia I.
; TITLE OF INVENTION: GENES EXPRESSED IN PROSTATE CANCER
; FILE REFERENCE: PA-0027-1 US
; CURRENT APPLICATION NUMBER: US/10/252,157
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/295,048
; PRIOR FILING DATE: 2001-05-31
; NUMBER OF SEQ ID NOS: 501
; SOFTWARE: PERL Program
; SEQ ID NO 421
; LENGTH: 1889
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20030190640A1 332919.4
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 1655-1681, 1892
; OTHER INFORMATION: a, t, c, g, or other
US-10-252-157-421

```

	Query Match	7.0%;	Score 144.4;	DB 12;	Length 1889;
	Best Local Similarity	99.3%;	Pred. No. 2.6e-26;		
	Matches 145;	Conservative	0;	Mismatches 1;	Indels 0; Gaps 0;
Qy	1929	GCGGGAGTGTTACTTCGGTTCCTCCAGGTTGGAGATTATCTCACCGGCCCCAGCTATATA	1988		
Db	5	GAGGGAGTGTTACTTCGGTTCCTCCAGGTTGGAGATTATCTCACCGGCCCCAGCTATATA	64		
Qy	1989	AGCTGACCGGTGTGTGGAGGGGCCACGACGGGCCCAACTCAGGGATTCTTCCACGACAGAA	2048		
Db	65	AGCTGACCGGTGTGTGGAGGGGCCACGACGGGCCCAACTCAGGGATTCTTCCACGACAGAA	124		
Qy	2049	AAACATACAAGACTCTTTCAGCCAAAC	2074		
Db	125	AAACATACAAGACTCTTTCAGCCAAAC	150		

RESULT 6

RESOLUTION 6
US-10-084-817-217
; Sequence 217, Application US/10084817
; Publication No. US20030119009A1
; GENERAL INFORMATION:
; APPLICANT: Susan Stuart
; APPLICANT: Jed G. Nuchtern
; APPLICANT: Sharon E. Plon
; APPLICANT: Jason M. Shohet
; TITLE OF INVENTION: GENES REGULATED BY MYCN ACTIVATION
; FILE REFERENCE: PA-0046 US

```

: CURRENT APPLICATION NUMBER: US/10/084,817
:
: CURRENT FILING DATE: 2002-02-25
: PRIOR APPLICATION NUMBER: 60/270,784
: PRIOR FILING DATE: 2001-02-23
: NUMBER OF SEQ ID NOS: 365
: SOFTWARE: PERL Program
: SEQ ID NO 217
: LENGTH: 1325
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc feature
: OTHER INFORMATION: Incyte ID No. US20030
: NAME/KEY: unsure
: LOCATION: 1258
: OTHER INFORMATION: a, t, c, g, or other
: US-10-084-817-217

```

Query Match	Score 6.8%	DB 14.14	Length 1325
Best Local Similarity 99.3%	Pred. NO. 1.3e-25		
Matches 142	Conservative 0	Mismatches 1	Indels 0
Gaps 0			
Qy 1932	GGAGTGTTACTTCGCTTCCAGGTTCCGAGGATTATCTCACCCGGCCCCCAGCTATATAAGC	1991	
Db 1	GGAGTGTTACTTCGCTTCCAGGTTCCGAGGATTATGTACCCGGCCCCCAGCTATATAAGC	60	
Qy 1992	TGACCGGTGTGGAGGGGCCAGCAGGGGCCAACTCCACGGGATTCCTTCCACGACAGAAAAA	2051	
Db 61	TGACCGGTGTGGAGGGGCCAGCAGGGGCCAACTCCACGGGATTCCTTCCACGACAGAAAAA	120	
Qy 2052	CATACAAGACTCCTTCAGCCAAC	2074	
Db 121	CATACAAGACTCCTTCAGCCAAC	143	

RESULT 7

US-09-923-876-5218
; Sequence 5218, Application US/09923876
; Patent No. US20020013958A1
; GENERAL INFORMATION:
; APPLICANT: Lalgudi, Raghunath V.
; APPLICANT: Kamigaki, Laura Y. (Ito)
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN SEEDLING
; FILE REFERENCE: PL-0012-1 CON
; CURRENT APPLICATION NUMBER: US/09/923,876
; CURRENT FILING DATE: 2001-08-06
; PRIOR APPLICATION NUMBER: 09/298,329
; PRIOR FILING DATE: 1999-04-21
; PRIOR APPLICATION NUMBER: 60/085,331
; PRIOR FILING DATE: 1998-05-05
; NUMBER OF SEQ ID NOS: 6332
; SOFTWARE: PERL Program
; SEQ ID NO 5218
; LENGTH: 184
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020013958A1 700456617H1
; NAME/KEY: unsure
; LOCATION: 67, 74, 98, 106, 122, 129
; OTHER INFORMATION: a, t, c, g, or other
US-09-923-876-5218

Query Match	5.2%	Score 129.4	DB 9	Length 184
Best Local Similarity	95.3%	Pred. No. 4.6e-23		
Matches 141	Conservative 0	Mismatches 6	Indels 1	Gaps 1
QV 1928	GGCGGGAGTGTACTTCGGTTCCAGGTTCCGAAGATTATCTCACCCGGCCCCAGCTATAT	1987		
qb 1	GGCGGGAGTGTACTTCGGTTCCAGGTTTGAAGATTATCTCACCCGGCCCCAGCTATAT	60		

QY 1988 AACGTC-ACGGTGTGGAGGGCCCGAGGCGCAACTCCAGGGATTCTTCCACGACAG 2046
|||||
Db 61 AAGCTGNACGGTNTGGAGGGCCCGAGGCGCAACCCAGGGANTCCTTCCACGACAG 120
|||||
QY 2047 AAAAAATACAAGACTCCTTCACGCAAC 2074
|||||
Db 121 ANAATATNCAAGACTCCTTCACGCAAC 148
|||||

RESULT 8

US-09-764-860-1029/c
; Sequence 1029, Application US/09764860
; Patent No. US20020094953A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008
; CURRENT APPLICATION NUMBER: US/09/764,860
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1198
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1029
; LENGTH: 6040
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-860-1029

Query Match 4.6%; Score 95.4; DB 9; Length 6040;
Best Local Similarity 72.8%; Pred. No. 2.1e-13;
Matches 123; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 1136 ATATATAGTATTGTGGTAGAGATGGGATTTGCCATGTTGCCAGGCTAGTATTGAAC 1195
|||||
Db 3417 ATTTTATATTTTATGTAGAGATGGGTTTCCATATGTCAGGCTGGTCTTGAAT 3358
|||||
QY 1196 CCTAAGCTAAGCAATCTCTGCTCTGCTCCCAAAATGTTGGGATTACAGGTGAAGC 1255
|||||
Db 3357 CCTGACCTCAGGCATCTGCTGCTCAGCCTCTCAATGTGCTGGGATTACAGCGGTGAGC 3298
|||||
QY 1256 CACTGACCGCTGATAGCTGGTTTCATTACTCTATTCTTGACCAC 1304
|||||
Db 3297 CACTGACCTGGCGGTTTTTTTCTTTCTTTTCTTTTGGAGAC 3249
|||||

RESULT 9

US-10-074-095-1029/c
; Sequence 1029, Application US/10074095
; Publication No. US2003007704A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC008C1
; CURRENT APPLICATION NUMBER: US/10/074,095
; CURRENT FILING DATE: 2002-02-14
; Prior application number: 09/764,860
; Prior filing date: 2001-01-17
; Prior application number: 60/179,065
; Prior filing date: 2000-01-31
; Prior application number: 60/180,628
; Prior filing date: 2000-02-04
; Prior application number: 60/214,886
; Prior filing date: 2000-06-28
; Prior application number: 60/217,487
; Prior filing date: 2000-07-11
; Prior application number: 60/225,758
; Prior filing date: 2000-08-14
; Prior application number: 60/220,963
; Prior filing date: 2000-07-26
; Prior application number: 60/217,496
; Prior filing date: 2000-07-11
; Prior application number: 60/225,447
; Prior filing date: 2000-08-14

; Prior application number: 60/218,290
; Prior filing date: 2000-07-14
; Prior application number: 60/225,757
; Prior filing date: 2000-08-14
; Prior application number: 60/226,868
; Prior filing date: 2000-08-22
; Prior application number: 60/216,647
; Prior filing date: 2000-07-07
; Prior application number: 60/225,267
; Prior filing date: 2000-08-14
; Prior application number: 60/216,880
; Prior filing date: 2000-07-07
; Prior application number: 60/225,270
; Prior filing date: 2000-08-14
; Prior application number: 60/251,869
; Prior filing date: 2000-12-08
; Prior application number: 60/235,834
; Prior filing date: 2000-09-27
; Prior application number: 60/234,274
; Prior filing date: 2000-09-21
; Prior application number: 60/234,223
; Prior filing date: 2000-09-21
; Prior application number: 60/228,924
; Prior filing date: 2000-08-30
; Prior application number: 60/224,518
; Prior filing date: 2000-08-14
; Prior application number: 60/236,369
; Prior filing date: 2000-09-29
; Prior application number: 60/224,519
; Prior filing date: 2000-08-14
; Prior application number: 60/220,964
; Prior filing date: 2000-07-26
; Prior application number: 60/241,809
; Prior filing date: 2000-10-20
; Prior application number: 60/249,299
; Prior filing date: 2000-11-17
; Prior application number: 60/236,327
; Prior filing date: 2000-09-29
; Prior application number: 60/241,785
; Prior filing date: 2000-10-20
; Prior application number: 60/244,617
; Prior filing date: 2000-11-01
; Prior application number: 60/225,268
; Prior filing date: 2000-08-14
; Prior application number: 60/236,368
; Prior filing date: 2000-09-29
; Prior application number: 60/251,856
; Prior filing date: 2000-12-08
; Prior application number: 60/251,868
; Prior filing date: 2000-12-08
; Prior application number: 60/229,344
; Prior filing date: 2000-09-01
; Prior application number: 60/234,997
; Prior filing date: 2000-09-25
; Prior application number: 60/229,343
; Prior filing date: 2000-09-01
; Prior application number: 60/229,345
; Prior filing date: 2000-09-01
; Prior application number: 60/229,287
; Prior filing date: 2000-09-01
; Prior application number: 60/229,513
; Prior filing date: 2000-09-05
; Prior application number: 60/231,413
; Prior filing date: 2000-09-08
; Prior application number: 60/229,509
; Prior filing date: 2000-09-05
; Prior application number: 60/236,367
; Prior filing date: 2000-09-29
; Prior application number: 60/237,039
; Prior filing date: 2000-10-02
; Prior application number: 60/237,038
; Prior filing date: 2000-10-02
; Prior application number: 60/236,370

;
; PRIOR FILING DATE: 2000-09-29
; PRIOR APPLICATION NUMBER: 60/236,802
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,037
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/237,040
; PRIOR FILING DATE: 2000-10-02
; PRIOR APPLICATION NUMBER: 60/240,960
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/239,935
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/239,937
; PRIOR FILING DATE: 2000-10-13
; PRIOR APPLICATION NUMBER: 60/241,787
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,474
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/246,532
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/249,216
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,210
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/226,681
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,759
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/225,213
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/227,182
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: 60/225,214
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/235,836
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/230,438
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/215,135
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 60/225,266
; PRIOR FILING DATE: 2000-08-14
; PRIOR APPLICATION NUMBER: 60/249,218
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,208
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,213
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,212
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,207
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,245
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,244
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,217
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,211
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,215
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,264
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,214
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/249,297
; PRIOR FILING DATE: 2000-11-17
; PRIOR APPLICATION NUMBER: 60/232,400
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/231,242
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/232,081
; PRIOR FILING DATE: 2000-09-08

;
; PRIOR APPLICATION NUMBER: 60/232,080
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,414
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/231,244
; PRIOR FILING DATE: 2000-09-08
; PRIOR APPLICATION NUMBER: 60/233,064
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/233,063
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,397
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,399
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/232,401
; PRIOR FILING DATE: 2000-09-14
; PRIOR APPLICATION NUMBER: 60/241,808
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,826
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,786
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/241,221
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/246,475
; PRIOR FILING DATE: 2000-11-08
; PRIOR APPLICATION NUMBER: 60/231,243
; PRIOR FILING DATE: 2000-09-08
;
; Query Match 4.6%; Score 95.4; DB 14; Length 6040;
; Best Local Similarity 72.8%; Pred. No. 2.1e-13;
; Matches 123; Conservative 0; Mismatches 46; Indels 0; Gaps 0;
;
QY 1136 ATATATAGTATTGTGGTAGAGATGGGATTTGCCATGTTGCCAGGCTAGTATTGAACT 1195
DB 3417 ATTTTATATATTTTAGTGGAGATGGGTTTCGCCATATTGCCAGGCTGGTCTTGAAC 3358
QY 1196 CCTAGCTAAGCAATCTTCTGCTCTGCTCCAAATGTTGGGATTACAGGTGAAGC 1255
DB 3357 CTGACCTCAGCAUCTGCTGCTCAGCTCTCATGTGCTGGGATTACAGCGGTGAGC 3298
QY 1256 CACTGACCCGGCTGATAGCTGGTTTCATTTACTTCTATTCTTTGACCAC 1304
DB 3297 CACTGACCTGGCCGTTTTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTTCTTTT 3249

RESULT 10

US-10-262-552-33
; Sequence 33, Application US/10262552
; Publication No. US20030125289A1
; GENERAL INFORMATION:
; APPLICANT: Gelb, Bruce D.
; APPLICANT: Tartaglia, Marco
; TITLE OF INVENTION: NOONAN SYNDROME GENE
; FILE REFERENCE: 2420/1J859-US1
; CURRENT APPLICATION NUMBER: US/10/262,552
; CURRENT FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: US 60/326,532
; PRIOR FILING DATE: 2001-10-01
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 300000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(300000)
; OTHER INFORMATION: where n may be a or g or c or t/u, unknown or other
US-10-262-552-33

Query Match 4.6%; Score 94.8; DB 14; Length 300000;
Best Local Similarity 77.1%; Pred. No. 3.3e-12;

	Matches	128;	Conservative	0;	Mismatches	37;	Indels	1;	Gaps	1;
Qy	1135	TATATATAGTATTTTGTGTAGAGATGGGATTTTGCATGTTCCCCAGGCTAGTATTGAAC	1194							
Db	130900	TAATTTTGTATTTTGTAGTACAGATGGGTTTTCACCATGTTGGCCAGGCTGTTTGAAC	130959							
Qy	1195	TCCTAAGCT-AAGCAATCTTCTGTCTGTGCTCCCAAAATGTTGGGATTACAGGTGTA	1253							
Db	130960	TCCCTGACTTCAGGTGATCTACTGTGCTCAGCCTCCAAAGTGATGGGATTACAGGTGTA	131019							
Qy	1254	GCCACTGCACCGGCTGATAGCTGGTTTCATTTACTCTATTCTTG	1299							
Db	131020	GCCACCGGAGCGGCCAAAACCTTTGTTTTTTTCTCTTTTGTG	131065							

```

RESULT 11
US-09-911-077A-19/c
; Sequence 19, Application US/09911077A
; Publication No. US20030114399A1
; GENERAL INFORMATION:
; APPLICANT: BLAKELY, RANDY D.
; APPLICANT: APPARSUNDARAM, SUBRAMANIAM
; APPLICANT: FERGUSON, SHAWN
; TITLE OF INVENTION: HUMAN AND MOUSE CHOLINE TRANSPORTER CDNA
; FILE REFERENCE: Velt:008US
; CURRENT APPLICATION NUMBER: US/09/911,077A
; CURRENT FILING DATE: 2001-07-23
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 119040
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (2347)..(90873)
; OTHER INFORMATION: N = A, C, G or T/U
US-09-911-077A-19

```

Query Match	4.5%;	Score 93.8;	DB 11;	Length 119040;
Best Local Similarity	66.7%;	Pred. No. 3.4e-12;		
Matches 134;	Conservative 0;	Mismatches 67;	Indels 0;	Gaps 0;
QY	1140	ATAGTATTTGTGTGAGAGATGGATTTTGCCTATGTTGCCACGGCTAGTATTGAATCTCTTA	1199	
Db	85255	ATTTGATTTTTTGTGAGAGATGGATTTTCCATGTTGCCAGGTGGCTTTGAATCTCTG		
QY	1200	AGCTAAGCAATCTTCTGTCTGTCTGCCAAAATGTTGGGATTAACAGTGTAGGCCACT	1259	
Db	85195	GCTCAACAACATCTCTGTCTCAGCGCTCCAAAAGTCTGGGATTGCAGGTGTGAGCCACT	85136	
QY	1260	GCACCCGGCTGATAGCTGGTTTCATTTACTCTATTTCTTTGACCACTCTGATCCATTGTGA	1319	
Db	85135	GTGCCCAGGCTACATACCTTGCTCTTGACCCCTTTTCCATATTTATTTATTTATTTT	85076	
QY	1320	AGTAAAAATGCTCCAATTATT	1340	
Db	85075	GAGACAGAGCGTCGCTCTGTT	85055	

```

RESULT 12
US-09-764-877-3933
; Sequence 3933, Application US/09764877
; Patent No. US20020147140A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC005
; CURRENT APPLICATION NUMBER: US/09/764,877
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 4031
; SOFTWARE: PatentIn Ver. 2.0

```

```

; SEQ ID NO 1933
;
; LENGTH: 32220
;
; TYPE: DNA
;
; ORGANISM: Homo sapiens
;
; FEATURE:
;
; NAME/KEY: SITE
;
; LOCATION: (9867)
;
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-877-3933

```

	Query Match	4.5%;	Score 93.2;	DB 10;	Length 32220;
	Best Local Similarity	79.2%;	Pred. No. 2.2e-12;		
	Matches 122;	Conservative 1;	Mismatches 30;	Indels 1;	Gaps 10;
Qy	1117	TCTCATTTTTTBTCAATATATATATAGATTGTTGGTAGAGATGGGATTTTTCACATGTG	1176		
Dd	30293	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGTGATTTTAGTAGAGATGGGTTTTTGCTATGTTG	30352		
Qy	1177	CCAGCGTAGTATTGAATCCTTAAGCT--AAGCAATCTTCTCGTGTCTCTGCCTCCCAAATG	1235		
Dd	30353	CCAGCGTGGTCTTGAATCTCTTAGCTCAAGTATCAGCCTTGCTCAGCCTCTCTAAAGTG	30412		
Qy	1236	TTGGGATTACAGGTGAAGCCACTGCACCCGGCT	1269		
Dd	30413	TTAGGATTATAGGTGTGAGCCACTGCACCTGGCT	30446		

```

RESULT 13
US-10-027-632-95562
; Sequence 95562, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 95562
; LENGTH: 555
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-95562

```

	Query Match	4.5%	Score 92.6	DB 12	Length 555
	Best Local Similarity	73.0%	Pred. No. 2.6e-13		
	Matches 119	Conservative 0	Mismatches 44	Indels 0	Gaps 0
QY	1135	TATATATAGTATTGTGGTAGAGATGGGATTTGCCATGTTGCCAGGCTAGTATTGCAAC	1194		
Db	320	TAATTTTGTATTTTGGTAGAGACAGGGTTTCACCATGTTGCCAGGCTGGTTTCGAAC	379		
QY	1195	TCTTAAGCTAAGCAATCTTCTGTGTCCTGCCCTCCCAAAATGTTGGGATTACACGGTGTAAG	1254		
Db	380	TCTTGAGCTCAGCAATACACATGCGCTCTGCGCTCCCAAGTGTTGGGATTACAGCGCGTGAG	439		
QY	1255	CAACTGCACCGGCTGATAGCTGGTTTCATTTACTCTATTCT	1297		

Db 440 TCACTGCATCTGCCAAATTTCTGTTTTTGTGCTTTTT 482

RESULT 14

US-10-027-632-95563
; Sequence 95563, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 10827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 95563
; LENGTH: 555
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-95563

Query Match 4.5%; Score 92.6; DB 12; Length 555;
Best Local Similarity 73.0%; Pred. No. 2.66-13;
Matches 119; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 1135 TATATAGTATTGTGGTAGAGATGGGATTTTGGCCATGTGCCAGGCTAGTATTGAAC 1194
Db 320 TAATTTTGTATTGTGGTAGAGACAGGGTTTCACCATGTGCCAGGCTGGTTTCGAAC 379
QY 1195 TCCTAAGCTAAGCAATCTTCTGCTCTGCTCTCCCAAAATGTTGGGATTACAGGTGAAG 1254
Db 380 TCCTGAGCTCAAGCAATACACTGCCTCTGCTCTCCCAAAAGTGTGGGATTACAGGCGTGAG 439
QY 1255 CCACTGCACCCGCGCTGATAGCTGGTTTCATTTACTCTATTCT 1297
Db 440 TCACTGCATCTGCCAAATTTCTGTTTTTGTGCTTTTT 482

RESULT 15

US-10-027-632-95562
; Sequence 95562, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 10827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363

; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 95562
; LENGTH: 555
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-95562

Query Match 4.5%; Score 92.6; DB 13; Length 555;
Best Local Similarity 73.0%; Pred. No. 2.6e-13;
Matches 119; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 1135 TATATAGTATTGTGGTAGAGATGGGATTTTGGCCATGTGCCAGGCTAGTATTGAAC 1194
Db 320 TAATTTTGTATTGTGGTAGAGACAGGGTTTCACCATGTGCCAGGCTGGTTTCGAAC 379
QY 1195 TCCTAAGCTAAGCAATCTTCTGCTCTGCTCTCCCAAAATGTTGGGATTACAGGTGAAG 1254
Db 380 TCCTGAGCTCAAGCAATACACTGCCTCTGCTCTCCCAAAAGTGTGGGATTACAGGCGTGAG 439
QY 1255 CCACTGCACCCGCGCTGATAGCTGGTTTCATTTACTCTATTCT 1297
Db 440 TCACTGCATCTGCCAAATTTCTGTTTTTGTGCTTTTT 482

Search completed: November 18, 2003, 07:42:07
Job time : 998.544 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 23:31:10 ; Search time 4326.29 Seconds
(without alignments)
11651.427 Million cell updates/sec

Title: US-10-005-337a-2

Perfect score: 2074

Sequence: 1 ctgcagcaagtacttaatg.....acaagactccttcagccaac 2074

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 22791392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gse_hum:*
18: em_gse_inv:*
19: em_gse_pin:*
20: em_gse_vrt:*
21: em_gse_fun:*
22: em_gse_mam:*
23: em_gse_mus:*
24: em_gse_pro:*
25: em_gse_rod:*
26: em_gse_pbg:*
27: em_gse_vrl:*
28: gb_gse1:*
29: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	461.2	22.2	522	28	AQ480395
2	260.6	12.6	586	28	AZ295534
3	241	11.6	735	9	AUI39209
4	181	8.7	371	13	BY091041

	5	158.4	7.6	353	13	BY079997
C 6	98.6	4.8	343	28	AQ584530	BY079997 BY079997
C 7	98.6	4.8	827	14	CD244827	AQ584530 RPCI-11-4
C 8	95.8	4.6	502	28	AQ207760	CD244827 AGENCOURT
C 9	95	4.6	412	9	AA494075	AQ207760 HS 3026_B
C 10	94.2	4.5	464	9	AA317468	AA494075 ng61c08_s
C 11	94.2	4.5	2772	11	BC038630	AA317468 QJ51fil.s
C 12	93.8	4.5	315	9	AI039619	BC038630 Homo sapi
C 13	93.6	4.5	369	10	BF738217	AI039619 ox28f05_x
C 14	93.6	4.5	464	10	BF829139	BF738217 CM3-KT003
C 15	93.6	4.5	723	28	AZ335882	BF829139 MRL1-HN007
C 16	93.2	4.5	488	10	BF415552	AZ335882 IM0065G12
C 17	93.2	4.5	532	29	BZ260574	BF415552 UI-R-CA1-
C 18	92.8	4.5	477	28	AZ406305	BZ260574 CH230-285
C 19	92.6	4.5	380	9	AA947380	AZ406305 IM0175C06
C 20	92.4	4.5	534	10	BF724783	AA947380 OD86H04_s
C 21	92.2	4.4	387	10	BF400140	BF724783 bx08f06_y
C 22	92.2	4.4	453	28	AQ187593	BF400140 UI-R-CA1-
C 23	92	4.4	563	12	BI438856	AQ187593 HS_3132_A
C 24	92	4.4	567	12	BI439182	BI438856 ic27e07_x
C 25	92	4.4	950	10	BG390793	BI439182 ic27e07_y
C 26	91.8	4.4	451	28	AQ108862	BG390793 602416521
C 27	91.6	4.4	419	9	AA468422	AQ108862 CIT-HSP-2
C 28	91.6	4.4	717	28	AQ343576	AA468422 nc78f02_x
C 29	91.4	4.4	442	28	AQ219900	AQ343576 RPC111-12
C 30	91.4	4.4	662	28	AZ365914	AQ219900 HS_3252_A
C 31	91.4	4.4	2073	11	AF289601	AZ365914 IM0112J19
C 32	91.2	4.4	470	9	AI814961	AF289601 Homo sapi
C 33	91	4.4	357	9	AW002875	AI814961 wx69b11_x
C 34	91	4.4	464	9	AA845825	AW002875 wt02a09_x
C 35	91	4.4	556	29	BZ606409	AA845825 ak85f12_s
C 36	91	4.4	587	9	AV720761	BZ606409 WHAAB93TR
C 37	91	4.4	589	9	AV720765	AV720761 AV720761
C 38	91	4.4	589	9	AV720514	AV720765 AV720765
C 39	91	4.4	639	28	AQ109988	AV720514 AV720514
C 40	91	4.4	988	10	BF791763	AQ109988 CIT-HSP-2
C 41	91	4.4	1017	12	BM472183	BF791763 602251956
C 42	90.8	4.4	284	14	T39841	BM472183 AGENCOURT
C 43	90.8	4.4	541	28	AQ895064	T39841 yallc09_r3
C 44	90.8	4.4	821	14	CD520493	AQ895064 HS_3175_A
C 45	90.6	4.4	430	9	AI288531	CD520493 AGENCOURT
						AI288531 ql91a04_x

ALIGNMENTS

RESULT 1
AQ480395/c
LOCUS
DEFINITION
RPIC-11-236B22-TV RPCI-11 Homo sapiens genomic clone RPCI-11-236B22
, genomic survey sequence.
ACCESSION
AQ480395
VERSION
AQ480395.1 GI:4662514
KEYWORDS
GSS.
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
Zhao, S., Adams, M.D., Nierman, W., Malek, J., de Jong, P. and Venter
J.C.
AUTHORS
Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready
Map Building
JOURNAL
Unpublished
COMMENT
Other GSSs: RPCI-11-236B22.TJ
Contact: Shaying Zhao, William Nierman, Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850
Tel: 301 838 0200
Fax: 301 838 0208
Email: hbe@tigr.org
Clones are derived from the human BAC library RPCI-11. For BAC

library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from Research Genet cs (info@resgen.com). BAC end search page: http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html.
Seq primer: T7
Class: BAC ends.

FEATURES

Location/Qualifiers
1. 522
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="GDB:7590285"
/db_xref="taxon:9606"
/clone="RPC1-11-236B22"
/sex="Male"
/cell_type="Lymphocytes"
/clone_lib="RPC1-11"
/notes="Vector: pBACe3.6; Site 1: EcoRI; Site 2: EcoRI;
RPC111 Human Male BAC Library"
168 a 100 c 113 g 141 t

BASE COUNT

Query Match 22.2%; Score 461.2; DB 28; Length 522;
Best Local Similarity 97.7%; Pred. No. 2.5e-19;
Matches 510; Conservative 1; Mismatches 5; Indels 6; Gaps 4;

```

QY 971 GTTACCAATAATCAATC-AGAATCCCTGGAGTGGGGC-CAGGATCTGTATTCTGAC 1028
DB 522 GTTACCAATAATCAATCAAGAATCCCTGGAGTGGGGCAGGATCTGTATTCTGAC 463
QY 1029 AGCTCCCAAGAGTATCTCTTCCCAACAGATTTGAGAACTTCAGCTCAATGACTAA 1088
DB 462 AGCTCCCAAGAGTATCTCTTCCCAACAGATTTGAGAACTTCAGCTCAATGACTAA 403
QY 1089 TCAGAGTCTGCCATTCCTAATATCTGCTCTCAATTTTBTATATATATATATATATT 1148
DB 402 TCAGAGTCTGCCATTCCTAATATCTGCTCTCAATTTTBTATATATATATATATATT 343
QY 1149 GTGGTAGAGTGGATTTTGGCCATTTGCCAGGTAGTATTGAATCTCTAAGCTAAGCA 1208
DB 342 TTGGTAGAGTGGATTTTGGCCATTTGCCAGGTAGTATTGAATCTCTAAGCTAAGCA 283
QY 1209 ATCTTCTGCTCTGCTCCCAAAATTTGGGATTAAGCTGTAAGCCACTGCACCCGGC 1268
DB 282 ATCTTCTGCTCTGCTCCCAAAATTTGGGATTAAGCTGTAAGCCACTGCACCCGGC 223
QY 1269 TGATAGTGGTTTCATTTACTCTATTTCTGACACTCTGATCCATTTTGAAGTAAAT 1328
DB 222 TGATAGTGGTTTCATTTACTCTATTTCTGACACTCTGATCCATTTTGAAGTAAAT 163
QY 1329 GCTCCAATTATTATGCTGTTTATAGAACAGGTAAAGCATGTCATGTGCTA---ATGCCAG 1385
DB 162 GCTCCAATTATTATGCTGTTTATAGAACAGGTAAAGCATGTCATGTGCTAATTATGCCAG 103
QY 1386 TGACATCATTAAGAAAGGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 1445
DB 102 TGACATCATTAAGAAAGGATTAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 43
QY 1446 GTGGCATGTATGGGGCTATTAG-CCAGACATCACTCCA 1486
DB 42 GTGGCATGTATGGGGCTATTAGCCCAAGACATCACTCCA 1

```

RESULT 2

AZ295534
LOCUS
DEFINITION
RPC1-23-105P3.TJ RPC1-23 Mus musculus genomic clone RPC1-23-105P3,
genomic survey sequence.
ACCESSION
AZ295534
VERSION
AZ295534.1 GI:9537319
KEYWORDS
GSS.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 586)
Zhao,S., Nierman,W., Feidblyum,T., Malek,J., Shatsman,S., Akinret
B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
Mouse BAC End Sequences from Library RPC1-23

TITLE JOURNAL COMMENT

Unpublished
Other_GSSs: RPC1-23-105P3.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPC1-23. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Plate: 105 row: P column: 3
Seq primer: SP6
Class: BAC ends.

FEATURES

Location/Qualifiers
1. 586
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPC1-23-105P3"
/sex="Female"
/lab_host="DH10B"
/clone_lib="RPC1-23"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:
EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
141 a 148 c 154 g 143 t

BASE COUNT ORIGIN

Query Match 12.6%; Score 260.6; DB 28; Length 586;
Best Local Similarity 75.9%; Pred. No. 1.8e-07;
Matches 403; Conservative 0; Mismatches 114; Indels 14; Gaps 6;
QY 1533 CTTCCCTCAGGCTGTTTACCAGGAATAGGATGTC---TGGGACAAAGTTTCCCTAA 1589
DB 21 CTTCCCTCAGGCTGATTATCCCAAGTAGGATGTCCCAAGCAACACTTCCAGCCAA 80
QY 1590 GTGAAGTGTGATAAGTCTGTCTATCAGAAAGATATTACTGGGGTGTGATATGTAGGC 1649
DB 81 CTGGAGTGTGATAAGTCTGTCTATCAGAAAGATATTACTGGGGTGTGATATGTAGGC 140
QY 1650 ATCTCATTTTCTTGTAGTAGTAGTCATATCAAGCTGACAAAGAA---AAAAAGGGCAGTG 1707
DB 141 --TTGCATTTTCTTGATAGTGTAGTCATTTATCTGACAAAGAAAGAAAGAGCAGCG 198
QY 1708 ATGTGTGCAATGTCAACAGAGCTGTCCCTGAC-TCTTGACAATAGGATGACTTGC 1766
DB 199 ATGTGTGCAATATTAAACAGGAGCTGTCCCTGCTTCCCGATACGTGGGACTCGC 258
QY 1767 ATTGTGAGCGATGTGATCACCACCAAGGAATGGCCCTCTCACATTTCTTCTGATTCA 1826
DB 259 ATTGTGAGCGGTGTGGTCACTGCCAAAGATGACCCCTCTCACATTTCTTCTGATTGC 318
QY 1827 CATATTTCAGAGGGTTAGCTTGTCTCCCTCCCTTTCAGCTTCCAGACTTCCAGACTGATCT 1886
DB 319 CATACGCCCGGC-----CAGCTTGTGTCATCTCCCTCTTGGGCTTCCAGACTTCT 373

```

QY 1887 GGAATGAAATTCACCTGCTCTAGTTGGCTCTTAATGGGGGGGAGGTACTTCGG 1946
Db 374 GGAATGAAATTCACCTGCTCTGAATTTGGCACTGGTGGGGCAGGGGTGACTTGGC 433
QY 1947 TTCCAGGTTGGAAGATTATCTCACCGGCCCGAGCTATATAAGCTGACCGGTGTGGAGG 2006
Db 434 TTCCAGGTTGGAAGATTATCTCACCGGCCCGAGCTATATAAGCTGACCGGTGTGGAGG 492
QY 2007 GGCCCAAGCGGCGCAACTCCAGGATTCCTTCCACGACGAGAAAAACATACA 2057
Db 493 GGCTCCACAGCGGCCAGTTCCAGGGGTTTCATCCACATGAGAGAAAAACATAGA 543

RESULT 3
LOCUS AU139209
DEFINITION AU139209 PLACE1 Homo sapiens cDNA clone PLACE1010155 5', mRNA
ACCESSION AU139209
VERSION AU139209.1 GI:11000730
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 735)
Ota.T., Nishikawa.T., Suzuki.Y., Ighii.S., Saito.K., Kawai.Y.,
Yamamoto.J., Wakamatsu.A., Nakamura.Y., Nagai.T., Sugano.S. and
Isogai.T.
HRI human cDNA project
Unpublished
Contact: Takao Isogai
Genomics Laboratory
Helix Research Institute
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
Tel: 81-438-52-3975
Fax: 81-438-52-3986
Email: genomics@hri.co.jp
HRI human cDNA project; 5'- & 3'-end one pass sequencing: Helix
Research Institute; cDNA library construction: Department of
Virology, Institute of Medical Science, University of Tokyo, and
Helix Research Institute.
Location/Qualifiers
1..735
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="PLACE1010155"
/tissue_type="placenta"
/clone_lib="PLACE1"
/notes="Vector: pME18SFL3"

BASE COUNT 240 a 154 c 181 g 156 t 4 others
ORIGIN

Query Match 11.6%; Score 241; DB 9; Length 735;
Best Local Similarity 100.0%; Pred. No. 2.2e-06;
Matches 241; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1834 ACAGAGGTAGCTGTGCTCCCTCCCTCTTACGCTTCCAGACACTGAGCTGGAATGA 1893
Db 10 ACAGAGGTAGCTGTGCTCCCTCCCTCTTACGCTTCCAGACACTGAGCTGGAATGA 69
QY 1894 AAATTCACCTGCTCTGAGTTGCTCTTAATGGGGGGGAGGTACTTCGTTCCCGAG 1953
Db 70 AAATTCACCTGCTCTGAGTTGCTCTTAATGGGGGGGAGGTACTTCGTTCCCGAG 129
QY 1954 GTTGAAGATTATCTCACCGGCCCGAGCTATATAAGCTGACCGGTGTGGAGGGCCCGAG 2013
Db 130 GTTGAAGATTATCTCACCGGCCCGAGCTATATAAGCTGACCGGTGTGGAGGGCCCGAG 189
QY 2014 CAGGGCCAACTCAGGGATTCCTTCCACGACGAGAAAAACATACAGACTCCTTCAGCCAA 2073
Db 190 CAGGGCCAACTCAGGGATTCCTTCCACGACGAGAAAAACATACAGACTCCTTCAGCCAA 249

```

QY 2074 C 2074

Db 250 C 250

RESULT 4

BY091041

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 371)

Okazaki.Y., Furuno.M., Kasukawa.T., Adachi.J., Bono.H., Kondo.S.,

Nikaido.I., Ootani.N., Saito.R., Suzuki.H., Yamanaka.I., Kiyosawa.H.,

Yagi.K., Tomaru.Y., Hasegawa.Y., Nogami.A., Schonbach.C.,

Cojibori.T., Baldarelli.R., Hill.D.P., Bult.C., Hume.D.A.,

Quackenbush.J., Schriml.L.M., Kanapin.A., Matsuda.H., Batalov.S.,

Beisel.K.W., Blake.J.A., Bradt.D., Brusic.V., Chothia.C., Corbani.

L.E., Cousins.S., Dalla.E., Dragani.T.A., Fletcher.C.F., Forrest

A., Frazer.K.S., Gaasterland.T., Gariboldi.M., Gissi.C., Godzik.A.

Gough.J., Grimond.S., Gustincich.S., Hirokawa.N., Jackson.I.J.,

Jarvis.E.D., Kanai.A., Kawaji.H., Kawasawa.Y., Kedzierski.R.M.,

King.B.D., Konagaya.A., Kurochkin.I.V., Lee.Y., Lenhard.B., Lyons

P.A., Maglott.D.R., Maltais.L., Marchionni.L., McKenzie.L., Miki

H., Nagashima.T., Numata.K., Okido.T., Pavan.W.J., Pertea.G.,

Pesole.G., Petrovsky.N., Pillai.R., Pontius.J.U., Qi.D.,

Ramachandran.S., Ravasi.T., Reed.J.C., Reed.D.J., Reid.J., Ring

B.Z., Ringwald.M., Sandelin.A., Schneider.C., Semple.C.A., Secou

M., Shimada.K., Sultana.R., Takenaka.Y., Taylor.M.S., Tesdale

R.D., Tomita.M., Verardo.R., Wagner.L., Wahlestedt.C., Wang.Y.,

Watanabe.Y., Wells.C., Wilming.L.G., Wynshaw-Boris.A., Yanagisawa

M., Yang.I., Yang.L., Yuan.Z., Zavalan.M., Zhu.Y., Zimmer.A.,

Carninci.P., Hayatsu.N., Hirozane-Kishikawa.T., Konno.H., Nakamura

M., Sakazume.N., Sato.K., Shiraki.T., Waki.K., Kawai.J., Aizawa.K.

Arakawa.T., Fukuda.S., Hara.A., Hashizume.W., Imotani.K., Ishii

Y., Itoh.M., Kagawa.I., Miyazaki.A., Sakai.K., Sakaki.D., Shibata

K., Shinagawa.A., Yasunishi.A., Yoshino.M., Waterston.R., Lander

E.S., Rogers.J., Birney.E. and Hayashizaki.Y.

Analysis of the mouse transcriptome based on functional annotation

Of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

12466851

PUBMED

COMMENT

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gsr.riken.go.jp

URL: http://genome.gsc.riken.go.jp/

Aizawa.K., Akimura.T., Arakawa.T., Carninci.P., Fukuda.S., Hirozane

T., Imotani.K., Ishii.Y., Itoh.M., Kawai.J., Konno.H., Miyazaki.A.

Murata.M., Nakamura.M., Nomura.K., Numazaki.R., Ohno.M., Sakai.K.

Sakazume.N., Sasaki.D., Sato.K., Shibata.K., Shiraki.T., Tagami

M., Waki.K., Watahiki.A., Muramatsu.M. and Hayashizaki.Y. Direct

Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

Genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system--384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

A cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES

source

Location/Qualifiers

1. .371

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="K630084E21"

/tissue type="heart"

/dev stage="10 days neonate"

/clone_lib="RIKEN full-length enriched, 10 days neonate heart"

BASE COUNT 91 a 94 c 105 g 81 t

ORIGIN

Query Match 8.7%; Score 181; DB 13; Length 371;

Best Local Similarity 81.3%; Pred. No. 0.012; Length 371;

Matches 248; Conservative 0; Mismatches 50; Indels 7; Gaps 3;

QY 1754 TAGGATGACTTGCATTCCTGAGCGATGTGATCACCACCAAGAAATGGCCCTCTCACATT 1813

Db 5 TGGATGACTCGGATTCCTGAGCGGTGGTCACTGCCAAGAAATGACCTCTCACATT 64

QY 1814 TCTTCTGATTCATATTACAGAGGTGATCTGCTCCCTCCCTCTTTCAGTCTCC 1873

Db 65 TCTTCTGATTCATATTACAGAGGTGATCTGCTCCCTCCCTCTTTCAGTCTCC 119

QY 1874 AGACATGAGTCGGAATGAATTCACCTGCTGAGTGGCTCTTAATGGGGCGG 1933

Db 120 AGACATGAGTCGGAATGAATTCACCTGCTGAGTGGCTCTTAATGGGGCGG 179

QY 1934 AGTGTACTTCCGTTCCAGGTTGGAAGATTATCTACCCGGCCCGAGCTATATAAGCTG 1993

Db 180 GGTGACTTGGCTTCCAGCTGGAAGATTATCTACCCAGCCCTAGCTATATAA-CGG 238

QY 1994 ACCGGTGTGAGGGGCCAGAGGCAACTCCAGGATTCCTTC-CAGCAGAGAAAC 2052

Db 239 GCTGTGTGAGGGGCTCCACAGGGCCAGTTCAGGGGTTTCATCCACAGAGAGAAAC 298

QY 2053 ATACA 2057

Db 299 ATAGA 303

RESULT 5

BY079997

LOCUS

DEFINITION BY079997 RIKEN full-length enriched, 10 days neonate heart Mus musculus cDNA clone K630022E19 5', mRNA sequence.

ACCESSION

BY079997

VERSION

BY079997.1

KEYWORDS

EST.

SOURCE

Mus musculus (house mouse)

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 353)

REFERENCE

AUTHORS

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gojobori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schriml, L. M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K. W., Blake, J. A., Bradt, D., Brusic, V., Chothia, C., Corbani, L. E., Cousins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest

A., Frazer, K. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedierski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Maitais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pertea, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J. U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Sempile, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Tesdale, R. D., Tomita, M., Vetraro, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wynshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E. and Hayashizaki, Y.

Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

Nature 420, 563-573 (2002)

22354683

12466851

Contact: Yoshihide Hayashizaki

Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center (GSC), Yokohama Institute

The Institute of Physical and Chemical Research (RIKEN)

1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan

Tel: 81-45-503-9222

Fax: 81-45-503-9216

Email: genome-res@gs.riken.go.jp,

URL: <http://genome.gsc.riken.go.jp/>

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane

, T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A.

, Murata, N., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K.

, Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami

, M., Waki, K., Watanabe, A., Muramatsu, M. and Hayashizaki, Y. Direct

Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with

Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)

Normalization and subtraction of cap-trapper-selected cDNAs to

prepare full-length cDNA libraries for rapid discovery of new

genes. Genome Res. 10 (10), 1617-1630 (2000)

RIKEN integrated sequence analysis (RISA) system-384-format

sequencing pipeline with 384 multicapillary sequencer. Genome Res.

10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA

encyclopedia: real-time sequence clustering for construction of a

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)

cDNA library was prepared and sequenced in Mouse Genome

Encyclopedia Project of Genome Exploration Research Group in Riken

Genomic Sciences Center and Genome Science Laboratory in RIKEN.

Division of Experimental Animal Research in Riken contributed to

prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for

further details.

FEATURES

source

Location/Qualifiers

1. .353

/organism="Mus musculus"

/mol_type="mRNA"

/strain="C57BL/6J"

/db_xref="taxon:10090"

/clone="K630022E19"

/tissue type="heart"

/dev stage="10 days neonate"

/clone_lib="RIKEN full-length enriched, 10 days neonate heart"

BASE COUNT 82 a 88 c 102 g 79 t

ORIGIN

Query Match 7.6%; Score 158.4; DB 13; Length 353;

Best Local Similarity 73.9%; Pred. No. 0.27;


```

DEFINITION      ol51f11.s1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:1527021.3' similar to SW:TS6 HUMAN P98066 TUMOR NECROSIS
FACTOR-INDUCIBLE PROTEIN TSG-6 PRECURSOR ;contains Alu repetitive
element.; mRNA sequence.
ACCESSION       AA917468.1 GI:3057358
VERSION         AA917468
KEYWORDS        Homo sapiens (human)
SOURCE          Homo sapiens
ORGANISM        Homo sapiens
REFERENCE       1 (bases 1 to 464)
AUTHORS        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
TITLE          Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
JOURNAL        NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
COMMENT        National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgabbs@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 417 Std Error: 0.00
Seq primer: -40ml3 fwd. BT from Amersham
High quality sequence stop: 345.
FEATURES       1..464
               Location/Qualifiers
               source
               1..464
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="IMAGE:1527021"
               /lab_host="DH10B"
               /clone_lib="Soares_NFL_T_GBC_S1"
               /notes="Organ: pooled; Vector: p77T3D-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Equal amounts of plasmid DNA from three normalized
libraries (fetal lung NbH19W, testis NHT, and B-cell
NCI-CGAP GCBI) were mixed, and ss circles were made in
vitro. Following HAP purification, this DNA was used as
tracer in a subtractive hybridization reaction. The driver
was PCR-amplified cDNAs from pools of 5,000 clones made
from the same 3 libraries. The pools consisted of
I.M.A.G.E. clones 297480-302087, 682632-687239,
26408-728711, and 729096-731399. Subtraction by Bento
Soares and M. Fatima Bonaldo."
BASE COUNT     118 a 101 c 96 g 149 t
ORIGIN
Query Match    4.5%; Score 94.2; DB 9; Length 464;
Best Local Similarity 70.4%; Pred. No. 1.4e+03;
Matches 126; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

QY 1117 TCTCATTTTBT CATATATATATAGTATTTCTGTGTAGAGATGGGATTTTGCATGTTG 1176
Db 1 TTTAATTTTAAATTTTAAATTTTAAATTTTTCATAGATGAGGCTTGCTGTATGTTG 60

QY 1177 CCAGGCTAGTATGAACCTCTAAGCTAAGCAATCTTCCTGTCTCTGCCCTCCCAAAATGT 1236
Db 61 CCAGGCTGTTGTTTAACTCCTCAGCTCAAGAGATCTCTGCCCTCTGCCCTCCAAAGTGA 120

QY 1237 TGGGATTACAGGTGAAGCCTGACCCGGCTGATAGCTGGTTTCATTTACTCTATTT 1295
Db 121 TGGGATTACAGGAGTAGCGCACTGCGCGGGTGAATTCCTGGTTTATGACTTGTCTTT 179

RESULT 11
BC038630
LOCUS
DEFINITION     BC038630 2772 bp mRNA linear HTC 04-MAR-2003
IMAGE:5263792, mRNA.
Homo sapiens, similar to hypothetical protein FLJ20489, clone
ACCESSION     BC038630
VERSION       BC038630
KEYWORDS      Homo sapiens (human)
SOURCE        Homo sapiens
ORGANISM      Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Strausberg, R.
Direct Submission
Submitted (15-OCT-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 73 Row: 1 Column: 4
This clone has the following problem: retained intron.
FEATURES       1..2772
               Location/Qualifiers
               source
               1..2772
               /organism="Homo sapiens"
               /mol_type="mRNA"
               /db_xref="taxon:9606"
               /clone="IMAGE:5263792"
               /cissue_type="Brain, hippocampus"
               /clone_lib="NIH MGC_95"
               /lab_host="DH10B"
               /lab_host="Vector: pBluescript"
               /note="Vector: pBluescript"
BASE COUNT     682 a 623 c 602 g 865 t
ORIGIN
Query Match    4.5%; Score 94.2; DB 11; Length 2772;
Best Local Similarity 68.9%; Pred. No. 4.8e+02;
Matches 144; Conservative 0; Mismatches 63; Indels 2; Gaps 1;

QY 1129 CATATATATATAGTATTTGTGTAGAGATGGGATTTTGCATGTTGCCAGGCTAGTA 1188
Db 997 CACCTATTTTATTCGATTTTATAGACAGCGGTTTGTGCCATGTTGCCAGGCTGGTC 1056

QY 1189 TTGAACCTCTAAGCTAAGCAATCTTCCTGTCTCTGCCCTCCCAAAATGTTGGGATTACAG 1248
Db 1057 TTCAACTCTGACCTCAGGTGAACCTTCTGCTCGGCTCCCAAGTGTGGGATTACAG 1116

QY 1249 TGTAAAGCACTGACCCGGCTGATA--GCTGGTTTCACTTACTCTATTTCTTGACCACTC 1306
Db 1117 CGTGAGCCACCAAGCCGCGGCTATAATTCAGTTTAAATAACACCATCTGAAGGCCACCA 1176

QY 1307 TGATCCATTTTGAAGTAAATAATGCTCCAA 1335
Db 1177 AGAGTCCAGTCCAGTAGGAATATTCCTCA 1205

RESULT 12
AI039619/c
LOCUS
DEFINITION     AI039619 315 bp mRNA linear EST 30-JUN-1998
IMAGE:4657865 3' similar to WP:12003.3 CE03672 ;, mRNA sequence.
ACCESSION     AI039619
VERSION       AI039619
KEYWORDS      Homo sapiens (human)
SOURCE        Homo sapiens
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 315)
REFERENCE

```

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps@mail.nih.gov
 This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 Seq primer: -40m13 fwd. ET from Amersham
 High quality sequence stop: 296.

FEATURES
 source
 1. .315
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:1657665"
 /dev_stage="8-9 weeks"
 /lab_host="DH10B"
 /clone_lib="Soares total fetus NB2HF8 9w"
 /notes="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site 1: Not I; Site 2: Eco RI; 1st strand cDNA was prepared from mRNA obtained from pooled 8-9 week (total) fetus material with a Not I - oligo(dT) primer [5' TGTACCAATCTGAAGTGGAGCGCGCTTAATTTTTTTTTTTT 3']. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 61 a 65 c 69 g 120 t
 ORIGIN

Query Match 4.5%; Score 93.8; DB 9; Length 315;
 Best Local Similarity 72.2%; Pred. No. 1.9e+03;
 Matches 122; Conservative 0; Mismatches 47; Indels 0; Gaps 0;
 QY 258 AAATAGTAATGACTCTGTTGCTTTCAGCAGGACATATCTATAAATAGGAGCTATACAAAG 317
 DB 177 AATACTAGCAGCGTGCTCGCTTCGCGACATATCTATAAATGGACGATACAG 118
 QY 318 AGATTAGCATGACCTCTGCGAAGATGACACAAATTTGGAACATTCATATATT 377
 DB 117 AAGATTAGCATGCGCCCTCGCGAAGATGACACCAAAATTCGTGAAGCGTTCCATATTA 58
 QY 378 AAAAATAAATAATAAAGAGAAAGGAAAAATTAAGAAATA 426
 DB 57 AAGAAAAAAGAAAGAAACAAACCCCAACAAACAAACAAAGA 9

RESULT 13
BF738217
LOCUS BF738217 369 bp mRNA linear EST 10-JAN-2001
DEFINITION CM3-KT0033-151200-572-f07 KT0033 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF738217
VERSION BF738217.1 GI:12064893
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 369)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE 20202663
PUBMED 10737800
COMMENT Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL.
 (<http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM3&t2=CM3-KT0033-151200-572-f07&t3=2000-12-15&t4=1>)
 Seq primer: puc 18 forward
 High quality sequence start: 15
 High quality sequence stop: 369.

FEATURES
 source
 1. .369
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="Adult"
 /clone_lib="KT0033"
 /notes="Organ: bladder tumor; Vector: puc18; Site 1: SmaI; Site 2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 90 a 85 c 78 g 116 t
 ORIGIN

Query Match 4.5%; Score 93.6; DB 10; Length 369;
 Best Local Similarity 67.3%; Pred. No. 1.8e+03;
 Matches 132; Conservative 0; Mismatches 64; Indels 0; Gaps 0;
 QY 1135 TATATAGTATTGTGTAGAGATGGGATTTTGCATGTTGCCAGGTAGTATTGAAC 1194
 DB 79 TAATTTTGTATTTTGGTAAAGATGGGTTTCCACCATTTTGGCCAGGCTGCTTGAAT 138
 QY 1195 TCCTAGCTAAGCAATCTTCCTGCTCTGCTCCCAAAATTTGGGATTACAGGTCTAG 1254
 DB 139 TCCTGACCAAGTATCTCCCTGCTCTGCTCCCAAGTGTGGGATTACAGGTGAA 198
 QY 1255 CCACTCCACCCGCTGATAGCTGTTTCATTACTCTATTCTTGACCCTCTGATCCAT 1314
 DB 199 TCATACACCCAGCCTTTAGATTGTTGTAAGTACATTTCTATGAGTTCACCAATGATGA 258
 QY 1315 TTGTAAAGTAAAAATGC 1330
 DB 259 AATTGCTTAAAGATGC 274

RESULT 14
BF829139/c
LOCUS BF829139 464 bp mRNA linear EST 13-JAN-2001
DEFINITION MR1-HN0070-151200-003-h04 HN0070 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF829139
VERSION BF829139.1 GI:12174344
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 464)
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 23:30:20 ; Search time 8578.08 Seconds
(without alignments)
11245.517 Million cell updates/sec

Title: US-10-005-337A-1
Perfect score: 2358
Sequence: 1 ggcattcttcattgtaaca.....caggctcgaggccaccatgg 2358

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_fun.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_hcg_hum.*
- 31: em_hcg_inv.*
- 32: em_hcg_other.*
- 33: em_hcg_mus.*
- 34: em_hcg_pln.*
- 35: em_hcg_rod.*
- 36: em_hcg_mam.*
- 37: em_hcg_vrt.*
- 38: em_sy.*
- 39: em_hgo_hum.*
- 40: em_hgo_mus.*
- 41: em_hgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2358	100.0	2358	6	AX468603	AX468603 Sequence
2	2301.8	97.6	142902	2	AC119234	AC119234 Mus muscu
3	879.8	37.3	229640	2	AC105469	AC105469 Rattus no
4	879.8	37.3	238344	2	AC097115	AC097115 Rattus no
c	804.8	34.1	838	11	BV062893	BV062893 S212P6025
5	646.2	27.4	723	10	AF478692	AF478692 Mus muscu
6	640.2	27.2	665	11	BV076484	BV076484 S212P6036
7	367.8	15.6	2074	6	AX468604	AX468604 Sequence
8	367.8	15.6	2074	6	AF131884	AF131884 Homo sapi
c	347.6	14.7	5011	9	AL590622	AL590622 Human DNA
9	128.6	5.5	1901	6	AX322775	AX322775 Sequence
10	128.6	5.5	1901	6	BD094076	BD094076 Shear str
11	128.6	5.5	1901	9	HGRNACINP	X83703 H.sapiens m
12	128.6	5.5	1901	11	G28603	G28603 human STS S
c	127.2	5.4	160350	2	AC074094	AC074094 Homo sapi
13	125.8	5.3	110480	10	AC122467	AC122467 Mus muscu
14	125.8	5.3	169126	2	AC132348	AC132348 Mus muscu
c	125.2	5.3	152346	2	AC102022	AC102022 Mus muscu
15	125.2	5.3	258445	2	AC122205	AC122205 Mus muscu
c	122.4	5.2	158357	9	AL365434	AL365434 Human DNA
16	121.6	5.2	1940	4	AF131883	AF131883 Oryctolog
c	119.6	5.1	25203	10	AL365324	AL365324 Mouse DNA
17	119.6	5.1	184754	2	AC022675	AC022675 Mus muscu
c	119.6	5.0	234469	2	AC119697	AC119697 Rattus no
18	117.8	5.0	153899	10	AL772285	AL772285 Mouse DNA
c	117.8	5.0	191485	2	AC091467	AC091467 Mus muscu
19	117.8	5.0	227165	10	AL626768	AL626768 Mouse DNA
c	117.8	5.0	243290	10	AL663088	AL663088 Mouse DNA
20	116.6	4.9	167591	2	AC136093	AC136093 Rattus no
c	116.6	4.9	179706	2	AC128045	AC128045 Rattus no
21	116.6	4.9	231573	2	AC123610	AC123610 Mus muscu
c	116.6	4.9	245489	2	AC105665	AC105665 Rattus no
22	116.6	4.9	267971	2	AC128995	AC128995 Rattus no
c	116.4	4.9	183205	2	AC133734	AC133734 Rattus no
23	116.4	4.9	238313	2	AC109696	AC109696 Rattus no
c	116.4	4.9	240006	2	AC094069	AC094069 Rattus no
24	116.2	4.9	249982	2	AC108661	AC108661 Rattus no
c	115.8	4.9	167223	2	AC113514	AC113514 Rattus no
25	115.8	4.9	219619	2	AC116128	AC116128 Mus muscu
c	115.4	4.9	244328	2	AC094938	AC094938 Rattus no
26	115.2	4.9	171403	10	AC124423	AC124423 Mus muscu
c	114.8	4.9	294703	2	AC103396	AC103396 Mus muscu
27	114.6	4.9	176926	2	AC138341	AC138341 Mus muscu
c	114.4	4.9	167601	2	AC115886	AC115886 Mus muscu

ALIGNMENTS

RESULT 1
AX468603
LOCUS AX468603 2358 bp DNA linear PAT 16-JUL-2002
DEFINITION Sequence 1 from Patent WO0246220.
ACCESSION AX468603
VERSION AX468603.1 GI:21901402
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Schwartz,B., Branellec,D. and Chien,K.
TITLE Sequences upstream of the carp gene, vectors containing them and
uses thereof

JOURNAL Patent: WO 0246220-A 1 13-JUN-2002;
Aventis Pharma S.A. (FR) ; The Regents of The University of
California at San Diego (US) ; Benoit, Patrick (FR)
FEATURES Location/Qualifiers
source 1..2358
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
BASE COUNT 636 a 567 c 542 g 613 t
ORIGIN
Query Match 100.0%; Score 2358; DB 6; Length 2358;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGATCCTTTTCATGTTTAAACAATATCAACCTTAACCCAGGGGAACAGCCTGCCTGACAGT 60
DB 1 GGATCCTTTTCATGTTTAAACAATATCAACCTTAACCCAGGGGAACAGCCTGCCTGACAGT 60
QY 61 GGCCTTGCACCCATGAATACTTCCTAGTCTAGTCCGTTTGTGAACCTCAGCCCATCCCA 120
DB 61 GGCCTTGCACCCATGAATACTTCCTAGTCTAGTCCGTTTGTGAACCTCAGCCCATCCCA 120
QY 121 ACACCTTCTGCAAGCCCATCTCTACAAGTGCTCATTCGGAATTTCTCTGGAGCTTCTCT 180
DB 121 ACACCTTCTGCAAGCCCATCTCTACAAGTGCTCATTCGGAATTTCTCTGGAGCTTCTCT 180
QY 181 TTCAGGATCAGCCTGATTTCTAGGGCAGCAGTTCTCAACCTGGGGGCTCGACCCCTTTGG 240
DB 181 TTCAGGATCAGCCTGATTTCTAGGGCAGCAGTTCTCAACCTGGGGGCTCGACCCCTTTGG 240
QY 241 GGGAAATCAACAGCCCTTTACAGGGTGCATATCATCTATCTCTATATGTCAAGTATTTA 300
DB 241 GGGAAATCAACAGCCCTTTACAGGGTGCATATCATCTATCTCTATATGTCAAGTATTTA 300
QY 301 CATTACGATTCGTAACAGTAGCAAAATACAGGTATGAAATAGCAATGAAATATTTAT 360
DB 301 CATTACGATTCGTAACAGTAGCAAAATACAGGTATGAAATAGCAATGAAATATTTAT 360
QY 361 GATTGAAGGTCAACCAACATAGGCGCCACACTGTTCTAGAGAAAAATCACCTGGGTG 420
DB 361 GATTGAAGGTCAACCAACATAGGCGCCACACTGTTCTAGAGAAAAATCACCTGGGTG 420
QY 421 GGGAAAGGTTTGGGAAAGCTTCTGTCATCTTCATCTTCAATCTTCAAAGTATGTGTTACA 480
DB 421 GGGAAAGGTTTGGGAAAGCTTCTGTCATCTTCTCAATCTTCAAAGTATGTGTTACA 480
QY 481 GAAAGCTTTTCACTGCTGCTGGGGCTCTTAGTAAGTCTAGTAGGAACCTGTATGTAC 540
DB 481 GAAAGCTTTTCACTGCTGCTGGGGCTCTTAGTAAGTCTAGTAGGAACCTGTATGTAC 540
QY 541 CAGGCTGCTCTTTATGGGTGGAGCCCAAGACGCATCGTGGGTGGAGCGAAGACGCAACCT 600
DB 541 CAGGCTGCTCTTTATGGGTGGAGCCCAAGACGCATCGTGGGTGGAGCGAAGACGCAACCT 600
QY 601 CACCTTCTAGCTCTGCATCGATAGCAAGTAGCCCTAATGTTTCTGTGCTAGGTGTCTATCT 660
DB 601 CACCTTCTAGCTCTGCATCGATAGCAAGTAGCCCTAATGTTTCTGTGCTAGGTGTCTATCT 660
QY 661 CTGTGAATCGAGATCTTGGCTTGTCTGAATTTAGGGAGGCAAAATACTCAGAGATTTC 720
DB 661 CTGTGAATCGAGATCTTGGCTTGTCTGAATTTAGGGAGGCAAAATACTCAGAGATTTC 720
QY 721 AAGACTGCTCAGCAGCCCAAGAGTCTTCTCAAAGGAAAGGTCTCAACTCTCAGCCCCCCC 780
DB 721 AAGACTGCTCAGCAGCCCAAGAGTCTTCTCAAAGGAAAGGTCTCAACTCTCAGCCCCCCC 780
QY 781 TTAGCTCTAGCTCAGGCTTGGAAACAAACGGCCACAGCAATGAGAAAGCTGCCATAGCTG 840
DB 781 TTAGCTCTAGCTCAGGCTTGGAAACAAACGGCCACAGCAATGAGAAAGCTGCCATAGCTG 840
QY 841 CTTGTCACTTCAAGAGGTCAAGAAAAATAGTGTTAACCATGAAAAAGAGACGACCAACAG 900
DB 841 CTTGTCACTTCAAGAGGTCAAGAAAAATAGTGTTAACCATGAAAAAGAGACGACCAACAG 900

DB 841 CTTGTCACTTCAAGAGGTCAAGAAAAATAGTGTTAACCATGAAAAAGAGACGACCAACAG 900
QY 901 TTATCCATTGATAGCGTCTCAGGACAGATAGACAGAGAGAAACACTAGGAGAGGGGAACC 960
DB 901 TTATCCATTGATAGCGTCTCAGGACAGATAGACAGAGAGAAACACTAGGAGAGGGGAACC 960
QY 961 CACGAGGACAGAGTATTAGTGTGTTTTCAGGGCAATCTCTGTACTGAAGATTCT 1020
DB 961 CACGAGGACAGAGTATTAGTGTGTTTTCAGGGCAATCTCTGTACTGAAGATTCT 1020
QY 1021 AGAAACACAATTTGCTGTTGAACAGCTGAAGTGGGGTGGGGTCTTACCCTCATGTTCA 1080
DB 1021 AGAAACACAATTTGCTGTTGAACAGCTGAAGTGGGGTGGGGTCTTACCCTCATGTTCA 1080
QY 1081 TCGAAGGTGAGTAGGAGAGACAGATATATGATGGCCAGCATAAACACATACACAACA 1140
DB 1081 TCGAAGGTGAGTAGGAGAGACAGATATATGATGGCCAGCATAAACACATACACAACA 1140
QY 1141 CCCTAATTAAACACTTCCCTCTTCTACTGACACCCCTTCACTCTCTCTTTTCATAAAAA 1200
DB 1141 CCCTAATTAAACACTTCCCTCTTCTACTGACACCCCTTCACTCTCTCTTTTCATAAAAA 1200
QY 1201 TAAAAAAGTATTTTATGTGGCTCTTACGATAGAATCTTTCTCGAACTATAAAAAGATC 1260
DB 1201 TAAAAAAGTATTTTATGTGGCTCTTACGATAGAATCTTTCTCGAACTATAAAAAGATC 1260
QY 1261 TAAATATTTATATTTTTCACATTTTAAATATCTTAGCGATGACAAGCCAGAAACAAGTAT 1320
DB 1261 TAAATATTTATATTTTTCACATTTTAAATATCTTAGCGATGACAAGCCAGAAACAAGTAT 1320
QY 1321 TTTTGGCTCTCTCAACAGCAAAAGCTTGGGGCTTTTGTTCGGTGTAGGAATAGAAACA 1380
DB 1321 TTTTGGCTCTCTCAACAGCAAAAGCTTGGGGCTTTTGTTCGGTGTAGGAATAGAAACA 1380
QY 1381 CGAGAGCCCGGTGATCTAGGCAGATGCTCTATCATTTAGCCCATGAGTCTCAGCGCTCAG 1440
DB 1381 CGAGAGCCCGGTGATCTAGGCAGATGCTCTATCATTTAGCCCATGAGTCTCAGCGCTCAG 1440
QY 1441 AGCCACATTTTCTCGGGCTCTTTAAGCTTTTCCACAGCATTTGGGAACTTTTACTGAC 1500
DB 1441 AGCCACATTTTCTCGGGCTCTTTAAGCTTTTCCACAGCATTTGGGAACTTTTACTGAC 1500
QY 1501 AGCATCCAGTTGCTCTCTGCTAGAACTGACATCATCTCTGTGCATCCTCTCGG 1560
DB 1501 AGCATCCAGTTGCTCTCTGCTAGAACTGACATCATCTCTGTGCATCCTCTCGG 1560
QY 1561 CCGGTTTTGGGTAGATCTCTGATTAGCCTTTAGAACTAGACCGGTGAGCCTGTGT 1620
DB 1561 CCGGTTTTGGGTAGATCTCTGATTAGCCTTTAGAACTAGACCGGTGAGCCTGTGT 1620
QY 1621 GCACATAATTATGGCCAGTGACACCATAGAGTCAAAAGTGAATTAAGTCTTTCAATTT 1680
DB 1621 GCACATAATTATGGCCAGTGACACCATAGAGTCAAAAGTGAATTAAGTCTTTCAATTT 1680
QY 1681 CTCCTAATGCTCGTACGATGGCATGTGTCAGGGCCATTTTGTGTCAGACATCCTCCAG 1740
DB 1681 CTCCTAATGCTCGTACGATGGCATGTGTCAGGGCCATTTTGTGTCAGACATCCTCCAG 1740
QY 1741 AGAATTTCAAAACAGATAGACAAAGTGGCACCCAGACCCCATCTCTTCCCTCGGGCTGA 1800
DB 1741 AGAATTTCAAAACAGATAGACAAAGTGGCACCCAGACCCCATCTCTTCCCTCGGGCTGA 1800
QY 1801 TTATCCCAAGAAATAGGATGTCCCAAGCAACACTTCCAGGCCAAGTGGAGTGTGATAA 1860
DB 1801 TTATCCCAAGAAATAGGATGTCCCAAGCAACACTTCCAGGCCAAGTGGAGTGTGATAA 1860
QY 1861 GTCCAGTTTATCAGAAGATATGCTGTAGTGTGATGACAGTGTGCTTTCTTGAT 1920
DB 1861 GTCCAGTTTATCAGAAGATATGCTGTAGTGTGATGACAGTGTGCTTTCTTGAT 1920
QY 1921 ACCTTAGTCTATAGAGCTGACAAAGAGGAAAAAGAGCAGCGATGTGGTGCAATATTA 1980
DB 1921 ACCTTAGTCTATAGAGCTGACAAAGAGGAAAAAGAGCAGCGATGTGGTGCAATATTA 1980

QY 1981 ACAGGAGCTGTCCTCCCTGCTTCCCGATACGTGGGATGATCGCATGCTGAGCGGTGG 2040
 |||||
 Db 1981 ACAGGAGCTGTCCTCCCTGCTTCCCGATACGTGGGATGATCGCATGCTGAGCGGTGG 2040
 |||||
 QY 2041 GTCACTGCCAAGGAATGACCTCTCACATTTCTTCTGATTCGATACGCCGCCGCCAG 2100
 |||||
 Db 2041 GTCACTGCCAAGGAATGACCTCTCACATTTCTTCTGATTCGATACGCCGCCGCCAG 2100
 |||||
 QY 2101 CTGTGTCATCTCCCTCTTGTGGCTTCCCGACACACTAAGCTCTGGAATGAAATTCACCTGCCT 2160
 |||||
 Db 2101 CTGTGTCATCTCCCTCTTGTGGCTTCCCGACACACTAAGCTCTGGAATGAAATTCACCTGCCT 2160
 |||||
 QY 2161 CTGAATTTGCCACTGTGTGGGGCAGGGGTGTGACTTTGCTTCCAGGCTTGGGAAGATTATC 2220
 |||||
 Db 2161 CTGAATTTGCCACTGTGTGGGGCAGGGGTGTGACTTTGCTTCCAGGCTTGGGAAGATTATC 2220
 |||||
 QY 2221 TCACCCAGCCCTAGCTATATACGGCTGTGTGGAGGGCTCCACAGGCCAGTCCAG 2280
 |||||
 Db 2221 TCACCCAGCCCTAGCTATATACGGCTGTGTGGAGGGCTCCACAGGCCAGTCCAG 2280
 |||||
 QY 2281 GGGTTTCATCCACAAGAGAGAAAAACATAGACTCGAGGCTTAGGAGCTTGCATGCTGCA 2340
 |||||
 Db 2281 GGGTTTCATCCACAAGAGAGAAAAACATAGACTCGAGGCTTAGGAGCTTGCATGCTGCA 2340
 |||||
 QY 2341 GTCGGAGGCCACCATGG 2358
 |||||
 Db 2341 GTCGGAGGCCACCATGG 2358
 |||||

RESULT 2.

AC119234
 LOCUS Mus musculus clone RP24-211P24, 142902 bp DNA linear HTG 11-JUN-2003
 DEFINITION Mus musculus clone RP24-211P24, *** SEQUENCING IN PROGRESS ***, 2
 ordered pieces.

AC119234
 AC119234.9 GI:31581760

HTG; HTGS PHASE2; HTGS_FULLTOP; HTGS_ACTIVEFIN.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 142902)

Birren,B., Nusbaum,C. and Lander,E.

Mus musculus, clone RP24-211P24

Unpublished

2 (bases 1 to 142902)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
 Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L.,
 Boukhgalter,B., Brown,A., Camarata,J., Campopiano,A., Chang,J.,
 Chazaro,B., Choepel,Y., Colangelo,M., Collins,S., Collymore,A.,
 Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
 Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
 Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
 Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., LaRoque,K., Lamazares,R.,
 Landers,T., Lehoczy,J., Levine,R., Lindblad-Toh,K., Liu,G.,
 MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
 McCarthy,M., McEwan,P., McKernan,K., Meldrim,J., Meneus,L.,
 Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R.,
 Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
 Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
 Raymond,C., Retton,K., Rieback,M., Riley,R., Riese,C., Rogov,P.,
 Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
 Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
 Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
 Topham,K., Travers,M., Travis,N., Triglio,J., Vassiliev,H.,
 Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
 Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

Direct Submission

Submitted (25-APR-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 142902)

AUTHORS

Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
 Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
 Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
 Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
 Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
 Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
 Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
 Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
 Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
 Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
 Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
 Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
 Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
 O'Neill,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
 Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Riese,C., Rogov,P.,
 Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
 Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
 Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Wu,X.,
 Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
 Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

TITLE

JOURNAL

COMMENT

Submitted (11-JUN-2003) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jun 11, 2003 this sequence version replaced gi:31455706.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center project name: L25236

Center clone name: 211_P_24

* NOTE: This is a 'working draft' sequence. It currently

* consists of 2 contigs. Gaps between the contigs

* are represented as runs of N. The order of the pieces

* is believed to be correct as given, however the sizes

* of the gaps between them are based on estimates that have

* provided by the submitter.

* This sequence will be replaced

* by the finished sequence as soon as it is available and

* the accession number will be preserved.

* 1 61126: contig of 61126 bp in length

* 61127 61226: gap of 100 bp

* 61227 142902: contig of 81676 bp in length.

Location/Qualifiers

1..142902

/organism="Mus musculus"

/mol_type="genomic DNA"

/db_xref="taxon:10090"

/clone="RP24-211P24"

/clone_lib="RPCI-24 Male Mouse BAC"

BASE COUNT 41376 a 30899 c 29987 g 40473 t 167 others

ORIGIN

Query Match

Best Local Similarity 97.6%; Score 2301.8; DB 2; Length 142902;

Matches 2314; Conservative 0; Mismatches 2; Indels 1; Gaps 1;

QY

1 GGATCCTTTTCATGTTTAAACAATATCAACCCCAAGGGAACAGCTGCTGACACT 60

Db

89306 GGATCCTTTTCATGTTTAAACAATATCAACCCCAAGGGAACAGCTGCTGACACT 89365

QY

61 GGCTTTGGCACCACCAATGATCTCTCTAGTCTAGTCCGTTTGTGAAACTCAGCCATCCCA 120

Db

89366 GGCTTTGGCACCACCAATGATCTCTCTAGTCTAGTCCGTTTGTGAAACTCAGCCATCCCA 89425

QY

121 ACATCTTTCGAAGCCCATCTCTCAAGGTCCTCATTTGGGAATTTCTTGGAGCTTCTCT 180

Db

89426 ACATCTTTCGAAGCCCATCTCTCAAGGTCCTCATTTGGGAATTTCTTGGAGCTTCTCT 89485

QY	181	TTTCAGGATCAGCCTGATTCTAGGCGAGCAGTTCTCAACCTGGGGGCTCGACCCCTTTCG	240	1261	TAAATATTATTATATTTTTCACATTTTAAATATCTTAGCGATGACAGCCAGAAACAAGTATT	1320	
Db	89486	TTTCAGGATCAGCCTGATTCTAGGCGAGCAGTTCTCAACCTGGGGGCTCGACCCCTTTCG	89545	Db	90566	TAAATATTATTATATTTTTCACATTTTAAATATCTTAGCGATGACAGCCAGAAACAAGTATT	90624
QY	241	GGGAATCAAAAGCACCCTTTACAGGGGTCAATATCATCTATCTATATGTACAGTATTTA	300	QY	1321	TTTTGGCCTCTCTCAACAGCAAAAGCTTTGGGGCCTTTTGTTCCTGTGTAGGAATAGAAACA	1380
Db	89546	GGGAATCAAAAGCACCCTTTACAGGGGTCAATATCATCTATCTATATGTACAGTATTTA	89605	Db	90625	TTTTGGCCTCTCTCAACAGCAAAAGCTTTGGGGCCTTTTGTTCCTGTGTAGGAATAGAAACA	90684
QY	301	CATTACGATTCTGTAACAGTAGCAAAATACAGGTATGAAATAGCAATAGAAATATTTTAT	360	QY	1381	CGAGAGCCCGGTGTATCTAGGCAGATGCTCTATCATCTAGCCCATGAGTCTCCAGCCTCAG	1440
Db	89606	CATTACGATTCTGTAACAGTAGCAAAATACAGGTATGAAATAGCAATAGAAATATTTTAT	89665	Db	90685	CGAGAGCCCGGTGTATCTAGGCAGATGCTCTATCATCTAGCCCATGAGTCTCCAGCCTCAG	90744
QY	361	GATTGAAGTCAACACAAATAGAGCGCGCCACACTGTTCTAGAGAAAATCACTGGGTG	420	QY	1441	ACGCACATTTTCTCGGGCTCTCTTAAAGCTTTTCCACAGCATTTGGGAACTTTTACTGAC	1500
Db	89666	GATTGAAGTCAACACAAATAGAGCGCGCCACACTGTTCTAGAGAAAATCACTGGGTG	89725	Db	90745	ACGCACATTTTCTCGGGCTCTCTTAAAGCTTTTCCACAGCATTTGGGAACTTTTACTGAC	90804
QY	421	GGGAAAGGTTTGGGAAAGCCTTTCTGTCATCTTCAATCTTCAAAAGTGTATGTTTCA	480	QY	1501	AGCATCCAAAGTTGTGCTTCTGCTAAGAACTGGACTCACATCTCTGTGTCATCACTTCGG	1560
Db	89726	GGGAAAGGTTTGGGAAAGCCTTTCTGTCATCTTCAATCTTCAAAAGTGTATGTTTCA	89785	Db	90805	AGCATCCAAAGTTGTGCTTCTGCTAAGAACTGGACTCACATCTCTGTGTCATCACTTCGG	90864
QY	481	GAAAGCCTTTACGCTGTTCTGCTGGGGCTCTTAGTAAGTCTGAGTAGGAACTGTATGTAC	540	QY	1561	CCCGTTTTGGGTAGATCTCTGTATAGCCTTCAGATTTAGAACACGGGTGAGCCTTGGT	1620
Db	89786	GAAAGCCTTTACGCTGTTCTGCTGGGGCTCTTAGTAAGTCTGAGTAGGAACTGTATGTAC	89845	Db	90865	CCCGTTTTGGGTAGATCTCTGTATAGCCTTCAGATTTAGAACACGGGTGAGCCTTGGT	90924
QY	541	CAGGCTGCTCTTATAGGTGGAGCCAGACGCATCGTGGGTGGAGCGGAGCGCAACCT	600	QY	1621	GCACTAAATTTATGCGCAGTGACACCATAGAGTCAAAAGTGCATTTACTGAATGCTTTCAATTT	1680
Db	89846	CAGGCTGCTCTTATAGGTGGAGCCAGACGCATCGTGGGTGGAGCGGAGCGCAACCT	89905	Db	90925	GCACTAAATTTATGCGCAGTGACACCATAGAGTCAAAAGTGCATTTACTGAATGCTTTCAATTT	90984
QY	601	CACCTTTAGCTCTGCATCATAGCAAGTAGCCCTAATGTTCTGCTGTAGGTGTATCT	660	QY	1681	CTTCTAAATGCTGGTAGGATGTCACAGGGCCATTTTAGCTGCAGACATCACTCCAG	1740
Db	89906	CACCTTTAGCTCTGCATCATAGCAAGTAGCCCTAATGTTCTGCTGTAGGTGTATCT	89965	Db	90985	CTTCTAAATGCTGGTAGGATGTCACAGGGCCATTTTAGCTGCAGACATCACTCCAG	91044
QY	661	CTGTGAATCGAGTCTTGGCTTGTGTAATPAGGGGCAACAAATACTCAGAGATTCT	720	QY	1741	AGAAATCCAAACAGATAGAGACAAGTGGCACCCAGACCCATCTCTTCCCTCGGGCTGA	1800
Db	89966	CTGTGAATCGAGTCTTGGCTTGTGTAATPAGGGGCAACAAATACTCAGAGATTCT	90025	Db	91045	AGAAATCCAAACAGATAGAGACAAGTGGCACCCAGACCCATCTCTTCCCTCGGGCTGA	91104
QY	721	AGACTGCTCAGAGCCGAGTCTCTCTCAAGGAAGGTCTCACTCTCAGCCCCCCC	780	QY	1801	TTATCCCCAGAAATAGGATGTCCTCCAAAGCAACACTTTCCAGGCCAACTGAGGTGCTGATAA	1860
Db	90026	AGACTGCTCAGAGCCGAGTCTCTCTCAAGGAAGGTCTCACTCTCAGCCCCCCC	90085	Db	91105	TTATCCCCAGAAATAGGATGTCCTCCAAAGCAACACTTTCCAGGCCAACTGAGGTGCTGATAA	91164
QY	781	TTAGCTCTGAGTCAGGCTTGGAAACAAACGGCCACAGGAATGAGAAAGCTGCCATAGCTG	840	QY	1861	GTCCAGTTTATCAGAAAGATATGGCTGTAAAGTGTGATGCAAGTGTGTCATTTTCTTGAT	1920
Db	90086	TTAGCTCTGAGTCAGGCTTGGAAACAAACGGCCACAGGAATGAGAAAGCTGCCATAGCTG	90145	Db	91165	GTCCAGTTTATCAGAAAGATATGGCTGTAAAGTGTGATGCAAGTGTGTCATTTTCTTGAT	91224
QY	841	CTTGTCACTTCAAGAGGTCAAGAAATAGTGTAAACATGAAACGAGAAGCAACACAG	900	QY	1921	AGTTTAGTCATATAGAGCTGACAAAGAGGAAAAAGAGCAGCGATGCTGGTGAATATTA	1980
Db	90146	CTTGTCACTTCAAGAGGTCAAGAAATAGTGTAAACATGAAACGAGAAGCAACACAG	90205	Db	91225	AGTTTAGTCATATAGAGCTGACAAAGAGGAAAAAGAGCAGCGATGCTGGTGAATATTA	91284
QY	901	TTATCCATTGATAGCTCTCAGACAGATAGGACAGAGAAACACTAGGAGGGGAACC	960	QY	1981	ACAGCAGCTGTCCTCGCTTCCCGATACGTGGGATGACTCGCATTTGCTGAGCGGTGTG	2040
Db	90206	TTATCCATTGATAGCTCTCAGACAGATAGGACAGAGAAACACTAGGAGGGGAACC	90265	Db	91285	ACAGCAGCTGTCCTCGCTTCCCGATACGTGGGATGACTCGCATTTGCTGAGCGGTGTG	91344
QY	961	CACGAAGGCAAGGTATTAGTGTGTTTTCAGGGCAATGCTTGTACTGAAGATTCT	1020	QY	2041	GTCACTGCCAAAGGAATGACCCCTCTCACATTTTCTCTGATTCGCATACGCCCGGCCAG	2100
Db	90266	CACGAAGGCAAGGTATTAGTGTGTTTTCAGGGCAATGCTTGTACTGAAGATTCT	90325	Db	91345	GTCACTGCCAAAGGAATGACCCCTCTCACATTTTCTCTGATTCGCATACGCCCGGCCAG	91404
QY	1021	AGAAACACAATTTGCTGGTTGAACAGCTGAAGTGGGGTGGGGTCTTACCCCATGTTC	1080	QY	2101	CTTGTCACTCTCCCTCTTGGGGCTTCCAGACACTAAGTCTGGAATGAAAAATTCACCTGCCT	2160
Db	90326	AGAAACACAATTTGCTGGTTGAACAGCTGAAGTGGGGTGGGGTCTTACCCCATGTTC	90385	Db	91405	CTTGTCACTCTCCCTCTTGGGGCTTCCAGACACTAAGTCTGGAATGAAAAATTCACCTGCCT	91464
QY	1081	TGGAAGGCTGAGTGAGGAGACAGATATATGATGSCCAGCATAAACAAACATACACAA	1140	QY	2161	CTGAATTTGGCCACTGTTGGGGCAGGGGTGTCATTTGGCTTCCCGCTCCAGGCTGGAAGATTATC	2220
Db	90386	TGGAAGGCTGAGTGAGGAGACAGATATATGATGSCCAGCATAAACAAACATACACAA	90445	Db	91465	CTGAATTTGGCCACTGTTGGGGCAGGGGTGTCATTTGGCTTCCCGCTTCCAGGCTGGAAGATTATC	91524
QY	1141	CCCTAATTAACACTTCCCTCTTCTACTGACACCCCTTCACTCTCTCTTTTCAATAAAAA	1200	QY	2221	TCACCCAGCCCTAGCTTATAAAGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAG	2280
Db	90446	CCCTAATTAACACTTCCCTCTTCTACTGACACCCCTTCACTCTCTCTTTTCAATAAAAA	90505	Db	91525	TCACCCAGCCCTAGCTTATAAAGGGCTGGTGTGGAGGGGCTCCACAGGGCCAGTTCCAG	91584
QY	1201	TAAAAAAGTATTTATGTGGCTTTTACGATAGAAATTTTCTCGAACTATAAAAAGATC	1260	QY	2281	GGGTTTCATCCACAGAGAGAAAAAATACACTCGAGG	2317
Db	90506	TAAAAAAGTATTTATGTGGCTTTTACGATAGAAATTTTCTCGAACTATAAAAAGATC	90565	Db	91585	GGGTTTCATCCACAGAGAGAAAAAATACACTCGAGG	91621

RESULT 3

AC105469
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-140118, WORKING DRAFT SEQUENCE, 9
 AC105469 229640 bp DNA linear HTG 15-NOV-2002
 unordered pieces.
 AC105469 GI:25007338
 HTG; HTGS PHASE1; HTGS DRAFT; HTGS_FULLTOP.
 Rattus norvegicus (Norway rat)
 Rattus norvegicus
 ORGANISM
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.

REFERENCE

AUTHORS

1 (bases 1 to 229640)
 Muzny,D,Marie., Metzker,M.Lee., Abramson,S., Adams,C., Alder,J.,
 Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
 Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,D.,
 Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
 Biswal,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
 Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
 Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
 Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
 Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
 Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
 Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
 Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
 Egan,A., Escoto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
 Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
 Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
 Gebregorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
 Gunaratne,P., Haaland,W., Hamill,C., Hamilton,C., Hamilton,K.,
 Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
 Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogue,M.,
 Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
 Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
 Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
 Kowitz,C., Kratt,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
 Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
 Lorusuhea,L., Loulseghe,H., Lozado,R.J., Lu,X., Ma,J.,
 Maheshwari,M., Mahindartine,M., Mahmoud,M., Malloy,K., Mangum,A.,
 Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
 Mawhney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
 Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
 Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
 Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
 Nwaoketeme,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
 Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkuch,C.,
 Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
 Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
 Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,P.,
 Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
 Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
 Shetty,J., Shwartsbeyn,A., Sisson,I., Sitter,C.D., Smajls,D.,
 Sneed,A., Sodergren,B., Song,X.-Z., Sorelle,R., Sosa,J.,
 Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
 Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
 Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
 Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
 Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
 Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
 Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Zhao,S., Dunn,D., von
 Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
 Weinstock,G. and Gibbs,R.A.

Direct Submission

Unpublished

2 (bases 1 to 229640)

Worley,K.C.

Direct Submission

Submitted (09-JAN-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 229640)
 Rat Genome Sequencing Consortium.

TITLE
JOURNAL

COMMENT

Direct Submission
 Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Nov 15, 2002 this sequence version replaced gi:23101653.
 The sequence in this assembly is a combination of BAC based reads
 and whole genome shotgun sequencing reads assembled using Atlas
 (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
 in the feature table below represents a scaffold in the Atlas
 assembly (a 'contig-scaffold'). Within each contig-scaffold,
 individual sequence contigs are ordered and oriented, and separated
 by sized gaps filled with Ns to the estimated size. The sequence
 may extend beyond the ends of the clone and there may be sequence
 contigs within a contig-scaffold that consist entirely of whole
 genome shotgun sequence reads. Both end sequences and whole genome
 shotgun sequence only contigs will be indicated in the feature
 table.

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: http://www.hgsc.bcm.tmc.edu/
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GNAG
 Center clone name: CH230-140118
 ----- Summary Statistics
 Assembly program: Phrap; version 0.990129
 Consensus quality: 215126 bases at least Q40
 Consensus quality: 217436 bases at least Q30
 Consensus quality: 218866 bases at least Q20
 Estimated insert size: 214300; sum-of-contigs estimation
 Quality coverage: 10x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 9 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence.
 * as soon as it is available and the accession number will
 * be preserved.

Accession	Length	Quality	Notes
1	215467	contig of 215467 bp in length	
2	215468	215567: gap of unknown length	
3	215568	contig of 3215 bp in length	
4	218782	gap of unknown length	
5	218882	contig of 1477 bp in length	
6	220359	contig of unknown length	
7	220459	gap of unknown length	
8	220460	contig of 1155 bp in length	
9	221615	221714: gap of unknown length	
10	221715	contig of 1346 bp in length	
11	223061	223160: gap of unknown length	
12	223161	contig of 1611 bp in length	
13	224771	gap of unknown length	
14	224872	contig of 1015 bp in length	
15	225886	gap of unknown length	
16	225987	contig of 2194 bp in length	
17	228181	228280: gap of unknown length	
18	229640	contig of 1360 bp in length.	

Location/Qualifiers
 1..229640
 /organism="Rattus norvegicus"
 /mol_type="genomic DNA"
 /db_xref="taxon:10116"
 /clone="CH230-140118"
 1..1268
 /note="wgs end extension
 clone_end:T7"
 4765..7050
 /note="wgs_end extension
 clone_end:T7"
 complement(7251..8156)

FEATURES
 source
 misc_feature
 misc_feature
 misc_feature


```
/notes="clone_boundary
clone_end:T7
site:
end sequence:RWAFX57TJB"
complement (7252..8118)
/notes="clone_boundary
clone_end:T7
site:
end sequence:RWAFX57TJC"
213542..215467
/notes="wgs_end_extension
clone_end:T7"
215568..216791
/notes="wgs_end_extension
clone_end:T7"
BASE COUNT 61330 a 47195 c 48097 g 63316 t 9702 others
ORIGIN
Query Match 37.3%; Score 879.8; DB 2; Length 229640;
Best Local Similarity 82.3%; Pred. No. 1e-245;
Matches 1209; Conservative 0; Mismatches 177; Indels 83; Gaps 14;
QY 879 ATGAACAGAGAGACACACAGTTATCCATTGTAGCGTCTCAGGACAGATAGGACAGAG 938
DB 140789 ATGGAACAGAGAGGCGCAAGTTGCTTCGATAGTGTCTCAGGACAGCCAGGACAGAG 140848
QY 939 AGAACACTAGGAGAGGAGACCCACAGGACAGCAAGGTATTAGTGTCTGGTTTCAGGCG 998
DB 140849 ----CACTAGGAGAGAGAGACCCACAGAGG-----TATCAGTGTGCTGGTTTCCAGGCG 140899
QY 999 AATGCTCTGTACTGAAGATTCTAGAACACAAATTTGCTGTTGAAACAGCTGAAGTGGGGT 1058
DB 140900 AATGCTCTCAATCGAAGTTCTAGAACACAGTTTACTGATGAAGAGCTGAAGTGGGT 140959
QY 1059 GGGG----GTTCTTACCCTATGTTCAATGGAAGGAGTGTAGGAGAGACAGATATATGATG 1115
DB 140960 GGGGTGAAGAATAACAAACCCAGTTTCACTGAAGGTTAGGAGGAGGACGACATACGATG 141019
QY 1116 GCCAGATACAAACATACACACACCCCTAATTAACACTTCCCTCTTCTACTGACACCCC 1175
DB 141020 GCCAGCAT-----ACAAACATAACACCCCTAATTAATGCTTCCCTCTGCTACTGACACT 141074
QY 1176 CTTCACTCTCTCTTTTCATA-----AAAAATAAAAAAAGT 1210
DB 141075 CCCTTCACCTCTCTTTCAATAAATAACAAACAAACAAAAAATCAAAAAAATACT 141134
QY 1211 ATTTATGTCCTTACGATAGATCTTTCTCTGAACTATATAAGATCTAAATATTTA 1270
DB 141135 ATTTATACATGACTCTGAAGATAAAATCTTTTC-----CTATAAAAGAGACTAAATATTTA 141189
QY 1271 TATTTTTCACATTTAATATCTTAGCGATGACNAGCCAGAACAAAGTATTTTTCGCTCT 1330
DB 141190 TATTTTTCATGTTTAACTATCTCAGTGATGAAGCCAGAGACAAAGTA--TTCTGCTCT 141248
QY 1331 CTCACAGCAAAAGCTTGGGCGCTT--TTGTTTCCGTGTTAGGAATAGAACACAGAGGCC 1389
DB 141249 CTCACAGCAAAAGCTTGGGCGCTTGGCTTGGTGGTGGTGGATAGAAC----- 141299
QY 1390 CGTGATATAGGAGATGCTCTATCATATTAGCCCATGAGTCTCAGGCTCAGACGACAT 1449
DB 141300 CATGATATCAGGAGAT-----GCCTATGAATCTCAGGCTCAAGTGCATGT 141347
QY 1450 TTTCTCGGCTCTTTAAGCTTT--TCCACAGCATTTGGAAACTTTTACTGACAGCATCA 1508
DB 141348 TTCTCAGGCTCTCCAGGCTTTTCCCCACAGCATTTGGAAAGCTTTTACTGACAGCGTCAA 141407
QY 1509 AGTTGCTCTTCTGAAGAACTGGACTCACATCTCTCTGTGATCATCTTTCGGCCCGCTTT 1568
DB 141408 AGTCGTCCTTCTGCTAATAACTGGGCTCATATGCTCTCTGACCACTCTGACCAATTT 141467
QY 1569 GGGGTAGATCTCTGATGACCTTTCAGATTTAGAACACGGTGAAGCTGTGGTGCATAT 1628
DB 141468 GAGGTCAATATCCGATTAGCCTTCTGTTTGTAGAGCACA-----TGCCATGCACTAAT 141520
```

1629 TATGGCCAGTGACACCATAGTCAAAGTGCAATTACTGAATGCTTTCAATTTCTCCTAAT 1688
DB 141521 TATGGCCAGTGACACCATAGTCAAAGTGCAATTACTGAATGCTTTCAATTTCTCATAAT 141580
QY 1689 GCTGTACGATGGCATGTGCACAGGGCCATTTTAGCTGTCAGACATCACTCCAGAGAAATCC 1748
DB 141581 GATGTAAAGTGGCGTGTCTATGGGGCCATTTTAGCCCTGGACATCACTCCAGAGAAATCT 141640
QY 1749 AAACAGATAGACAAAGTGGGACCCAGACCCATCTCTTCCCTCGGGCTGATTATCCCC 1808
DB 141641 GAACAGATATAGCAAGTGTCCAAAGACCCA---CTTCCCCCGGGGTGTTTATTTCCC 141696
QY 1809 AGAAATAGGATGTCCAAAGCAACACTTCCAGCAGCAACTGGAGTGTGTATAGTCCAGTT 1868
DB 141697 AGAATAGGATGTCCAAAGCAACACTTCCAGCAGCAACTGGAGTGTGTATAGCCAGTT 141756
QY 1869 ATCAGAAAGATATGGCTGAAGTGTGATGCACAGTCTTTCGATTTTCTTGATACGTTAGT 1928
DB 141757 ATCAGAAAGATATTCGCGTGAGTGTGATGCACAAATGCTTTCGACTTTCCCTGATAGTTAGT 141816
QY 1929 CATATGAGAGCTGACAAAG 1988
DB 141817 CATACGAAAGCTGACAG 141876
QY 1989 CTGTCCTCTGGCTTCCCGATAGTGGGATGACTCGCATTTGCTGAGCGGTGTGGTCACTGC 2048
DB 141877 CTGT--CCCTGGCTTCCCGATAGTGGGATGACTCGCATTTGCTGAGCGGTGTGGTCACTGC 141935
QY 2049 CAAAGAAATGACCCCTCTCACTTTCTTCTGATTCGCATACGCGCGCCAGCTTGTCTAT 2108
DB 141936 CAATGGAATGGCCCTCTCACTTTCTTCTGATTCGCACACACACGCGCAGCTTGTCTAT 141995
QY 2109 CTCCTCTTGGCTTCCCGAGACACTAAGTCTCGAATGAAATTCACCTGCCTCTGAATTTG 2168
DB 141996 CTCCTCTTGGCTTCCCGAGACACTAAGTCTCGAATGAAATTCACCTGCCTCTGAATTTG 142055
QY 2169 GCACCTGTTGGGGGAGGGGTGTGACTTGGCTTCCAGGCTGGAAGATTATCTCACCCAG 2228
DB 142056 GCACAGCGGAGGAGGGGTGTGACTTGGCTTCCAGGCTGGAAGATTATCTCACCCAG 142115
QY 2229 CCTAGCTATATAACGGGCTGTGTGGAGGGCTCCACAGGCGCCAGTTCACAGGGGTTTAT 2288
DB 142116 TCCTAGCTATATAAGAGGCTGGGGTGGAGGGCTCCACAGGCGCCAGTTCACAGGGGTTTAT 142175
QY 2289 CCACAG 2317
DB 142176 CCACAG 142204

RESULT 4
AC097115 238344 bp DNA linear HTG 14-NOV-2002
LOCUS Rattus norvegicus clone CH230-26A2, *** SEQUENCING IN PROGRESS ***,
DEFINITION 2 unordered pieces.
ACCESSION AC097115
VERSION AC097115.6 GI:24956605
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE Rattus norvegicus (Norway rat)
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 238344)
Muzny, D.Marie., Metzker, M.Lee., Abramson, S., Adams, C., Alder, J.,
Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D.,
Anyalebechi, V., Ayodeji, A., Ayodeji, M., Baca, E., Baden, H.,
Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benahmed, F.,
Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M.,
Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E.,
Cardenas, V., Carter, K., Cavazos, I., Ceasar, H., Chen, R.,
Chacko, J., Chavez, D., Chen, R., Chen, Y., Chen, Z., Chu, J.,
Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L.,

Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D., Delgado,O., Denison,S., Deramo,C., Ding,Y., Dinh,H., Divya,K., Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Faves,K., Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G., Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P., Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M., Gebregorgis,E., Gest,K., Gill,R., Grady,M., Guerra,W., Guevara,W., Gunaratne,P., Haaland,M., Hamil,C., Hamilton,C., Hamilton,K.K., Harvey,Y., Havlik,P., Hawes,A., Henderson,N., Hernandez,J., Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M., Hollins,B., Howells,S., Huliyk,S., Hume,J., Idlebird,D., Jackson,A., Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Joliviet,A., Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C., Kowls,C., Kraft,C., Li,X., Lebow,H., Levay,J., Lewis,L., Li.Z., Liu,J., Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J., Lorensheva,L., Loulsegged,H., Lozado,R.J., Lu,X., Ma,J., Maheshwari,M., Mahindartine,M., Mahmoud M., Malloy,K., Mangum,A., Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E., Maxinney,S., McLeod,M.P., McNell,T.Z., Meenen,E., Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S., Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L., Nantervis,C., Neal,D., Newton,N., Nguyen.N., Norris.S., Nwaokemelehu,O., Okwuonu,G., Olarnpunsaagoon,A., Pal,S., Parks,K., Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C., Plopper,F., Polindexer,A., Popovic,D., Primus,E., Pu,L.-L., Puzo,M., Quiroz,J., Rachlin,B., Reeves,K., Regier,M.A., Reigh,R., Reilly,B., Reilly,M., Ren.Y., Reuter,M., Richards,S., Riggs,F., Rives,C., Rodkey.T., Rojas,A., Rose,M., Rose.R., Ruiz,S.J., Sanders,W., Savery,G., Scherer,S., Scott.G., Shatsman,S., Shen.H., Shetty,J., Shivartsbeyn,A., Sisdon,I., Sitter,C.D., Smajds,D., Snead,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J., Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C., Taylor,T., Thomas,N., Thomas.S., Tingey.A., Trejos,Z., Usmani,K., Valas,R., Vexa,V., Villasana,D., Waldron,L., Walker.B., Wang,J., Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F., Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K., Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon.V., Yu.F., Zhang,J., Zhou,J., Zhou.X., Zhao,S., Dunn,D., von Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O., Weinstock,G. and Gibbs,R.A.

Direct Submission
Unpublished

2 (bases 1 to 238344)
Worley,K.C.
Direct Submission
Submitted (11-OCT-2001) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 238344)
Rat Genome Sequencing Consortium.
Direct Submission
Submitted (14-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Nov 14, 2002 this sequence version replaced gi:22855482.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly ('contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----

QY	1390	CGTGATCTAGGAGATGCTCTATCATTAGCCCATGAGTCTCAGCCTCAGACGACATT	1449
Db	26929	CATGTATCTAGGAGAT-----GCCTATGAATCTCAGGCTCAAGTGATGTT	26976
QY	1450	TTTCTCGGGCTCTCTTAAGCTTT-TCCACAGCATTGGGAACTTTACTGACAGCATCCA	1508
Db	26977	TTTCTCAGGCTCTCCAGGCTTTCCCCACAGCATTGGAAGCTTTACTGACAGCGTCAA	27036
QY	1509	AGTTGTCTTCTGCTAAGAACTGGACTCACAATCTCTCTGTCGATCACTTCGGGCCCTTTT	1568
Db	27037	AGTCGTCTCTCTAATAACTGGGCTCATATGTCTCTGGACCACTCTGACCCATTCT	27096
QY	1569	GGGTAGATCCTCTGATTAGCCTTCAGATTAGAAACAGGTGAGCTGTGGTCACTAAT	1628
Db	27097	GAGGTCAATATTCGGATTAGCCTTCCTGTTTATAGCACA-----TGCCATGCACTAAT	27149
QY	1629	TATGGCCAGTGACACCATAGAGTCAAAAGTGCATTACTGAATGCTTTCATTTTCTCCTAAT	1688
Db	27150	TATGGCCAGTGACACCATAAAGTAAAGTGCATTACTGAATGCTTTCATTTTCTCATAAT	27209
QY	1689	GCTGTACAGTGCATGTGCAGGGCCATTTTATGCTGACATCACTCCAGAGAAATCC	1748
Db	27210	GATGGTAAAGTGGCGTGTATGGGGCCATTTTATGCTGACATCACTCCAGAGAAATCT	27269
QY	1749	AAACATAGACAGTGCACCCAGACCCCATCTCTCTCCCTCGGGCTGATTATCCCC	1808
Db	27270	GAACAGATATAGACAGTGTCAACCAAGACCCA---CTTCCCCCGGGCTGTTTATTCCT	27325
QY	1809	AGAAATAGATGTCCAAAGCAACACTTCCAGCCAACTGGAGTGTGATAGTCCAGTT	1868
Db	27326	AGGAATAGATGTCCAAAGCAACACTTCCAGCCAACTGGAGTGTGATAGTCCAGTT	27385
QY	1869	ATCAGAAAGATATGGTGAAGTGTGATGACAGTCTTGCATTTTCTTGATACGTTAGT	1928
Db	27386	ATCAGAAAGATATGGCGTGAAGTGTGATGACAAATGCTTGCACTTTTCTTGATAGTTAGT	27445
QY	1929	CATATCAGAGCTGACAAAGAGGAAAGAGAGCGATGCTGTCGAATATTAACAGCGAG	1988
Db	27446	CATACGAAAGCTGACAGAGAGAGAGAGGAGCGATGCTGTCGAATATTAACAGCGAG	27505
QY	1989	CTCTCCCTGGCTTCCCGATACCTGGGATGACCTGCGATTGCTGAGCGGTGTGCTCACTGC	2048
Db	27506	CTGT-CCCTGGCTTCCCGATAGTAGATGACTGCGATTGCTGGGCGGTGTGCTCACTGC	27564
QY	2049	CAAAGAAATGACCTCTCAGTTTCTTCTGATTCGCATACGCGCGGCGGCTGTCAT	2108
Db	27565	CAATGAATGGCCCTCTCAGTTTCTTCTGATTCGCATACGCGCGGCGGCTGTCAT	27624
QY	2109	CTCCCTTGGGCTTCCAGACACTAAGTCTGGAATGAAATTCACCTGCCTCTGAATTG	2168
Db	27625	CTCCCTTGGGCTTCCAGACACTAAGTCTGGAATGAAATTCACCTGCCTCTGAATTG	27684
QY	2169	GCCACTGGTGGGGCAGGGGTGTGACTTGGCTTCCAGGCTGGAAGATTATCTCACCCAG	2228
Db	27685	GCCACAGCGGAGCGAGGGGTGTGACTTGGCTTCCAGGCTGGAAGATTATCTCACCCAG	27744
QY	2229	CCTAGCTATATAACGGGTGTGTGCGAGGGGCTCCACAGGGCAGTTCAGGGGTTTAT	2288
Db	27745	TCCTAGCTATATAAGAGGTGGGGTGGAGGGGCTCCACAGGGGCTCCAGGGGTTTAT	27804
QY	2289	CCACAGAGAGAAACATAGACTCAGG	2317
Db	27805	CCACAGAGAGAGAAACATAGACTCAGG	27833

RESULT 5
BV062893/c
LOCUS S212P60252FC3.T0 CZECHII/Ei Mus musculus STS genomic, sequence
DEFINITION tagged site.
ACCESSION BV062893
VERSION BV062893.1 GI:31178688

KEYWORDS	STS.
SOURCE	Mus musculus (house mouse)
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 838)
AUTHORS	Wade, C.M., Kulbokas, E.J. III, Kirby, A.W., Zody, M.C., Mullikin, J.C., Lander, E.S., Lindblad-Toh, K. and Daly, M.J.
TITLE	The mosaic structure of variation in the laboratory mouse genome
JOURNAL	Nature 420 (6915), 574-578 (2002)
MEDLINE	22354684
PUBMED	12456852
COMMENT	Contact: Kerstin Lindblad-Toh Whitehead Institute for Biomedical Research, Center for Genome Research 320 Charles Street, Cambridge, MA 02141, USA Tel: 6172521477 Fax: 6172580903 Email: kersli@genome.wi.mit.edu Primer A: None Primer B: None STS size: 838 Protocol: WGS-discovery: Paired-end low-coverage whole genome shotgun reads were generated from 129S1/SvimJ, C3H/HeJ, and BALB/cByJ. The WGS reads were placed uniquely on the MGSCv3 C57BL/6J assembly and SNP detection was carried out by SSAHA-SNP. 225,000 reads were annotated as STSs and 81,000 SNPs were annotated with alleles from C57BL/6J and the strain from which the particular read came. The validation rate for these SNPs was estimated at approximately 98%. Location/Qualifiers 1. 838 /organism="Mus musculus" /mol_type="genomic DNA" /strain="CZECHII/Ei" /db_xref="taxon:10090" /map_pos="19 27-787 35750684-35751444" /clone_lib="CZECHII/Ei" <1. >838 STS BASE COUNT 235 a 164 c 208 g 231 t ORIGIN Query Match 34.1%; Score 804.8; DB 11; Length 838; Best Local Similarity 98.9%; Pred. No. 5.2e-224; Matches 831; Conservative 0; Mismatches 7; Indels 2; Gaps 2; QY 1004 CTGTACTGAAGCTTCTAGAAACACAAATTTGCTGTTGAACAGCTGAAGTGGGTGGGG 1063 Db 838 CTCGTTCTGAAGGTTCTAGAAACACAAATTTGCTGG-TGAACAGCTGAAGTGGGTGGGG 780 QY 1064 TTCTTACCCCATGTTTCATGGAAGGGTGAAGTGAAGAGAGACAGATATATATGATGCGCAGCAT 1123 Db 779 TTCTTACCCCATGTTTCATGGAAGGGTGAAGTGAAGAGAGACAGATATATATGATGCGCAGCAT 720 QY 1124 AACAAACATACACACACCCCTAATTAACACTTCCCTCTTCTACTGACACCCCTTCACTC 1183 Db 719 AACAAACATACACACACCCCTAATTAACACTTCCCTCTTCTACTGACACCCCTTCACTC 660 QY 1184 TCCTCTTTTCATAAAAATAAAAAAGTATTTTATGTGGCTCTTACGATAGAAATCTTTCTCT 1243 Db 659 TCCTCTTTTCATAAAAATAAAAAAGTATTTTATGTGGCTCTTACGATAGAAATCTTTCTCT 600 QY 1244 CGAACTATAAAAAGATCTAAATATTTATATTTTTCACATTTTAAATATCTTAGCGATGACA 1303 Db 599 CGAACTATAAAAAGATCTAAATATTTATATTTTTCACATTTTAAATATCTTAGCGATGACA 540 QY 1304 AGCCAGAAACAGTATTTTTCCTCTCTCAACAGCAAGCTTGGGGCTTTTGTTCCT 1363 Db 539 AGCCAGAAACAGTATTTTTCCTCTCTCAACAGCAAGCTTGGGGCTTTTGTTCCT 481 QY 1364 GTGTTAGGAATAGAACACGAGAGCCCGGTATCTAGGCAGATGCTCTATCATTAGCCCA 1423

Db 480 GTGTTAGGAATAGAACACGAGACCCCGTGTATCTAGGCAGATGCTCTATCATTTAGCCCA 421
QY 1424 TGAGTCTCAGAGCTCAGACGACATTTTCTCGGGCTCTCTTAAGCTTTTCCACAGCAT 1483
Db 420 TGAGTCTCAGAGCTCAGACGACATTTTCTCGGGCTCTCTTAAGCTTTTCCACAGCAT 361
QY 1484 TGGGAACTTTACTGACAGCATCCAAAGTTGTGCTTCTGCTTAAGAACTGGAGCTCAGATCTC 1543
Db 360 TGGGAACTTTACTGACAGCATCCAAAGTTGTGCTTCTGCTTAAGAACTGGAGCTCAGATCTC 301
QY 1544 TCTGTGCATCACTTCGGCCGGTTTGGGTAGATCTCTGATAGGCTTCAGATTTAGAA 1603
Db 300 TCTGTGCATCACTTCGGCCGGTTTGGGTAGATCTCTGATAGGCTTCAGATTTAGAA 241
QY 1604 CACGGTGAGCTGTGTGTGCACCTAAATATGGCCAGTGACACCATAGAGTCAAAAGTGCATTA 1663
Db 240 CACGGTGAGCTGTGTGTGCACCTAAATATGGCCAGTGACACCATAGAGTCAAAAGTGCATTA 181
QY 1664 CTGAATGCTTTCAATTTCTCCTAATCTCGTACGATGGCATGTCACAGGGCCATTTTACG 1723
Db 180 CTGAATGCTTTCAATTTCTCCTAATCTCGTACGATGGCATGTCACAGGGCCATTTTACG 121
QY 1724 TGCAGACATCACTCCAGAGAAATTCACACAGATAGAGACAAAGTGGCCACCCAGCCATCT 1783
Db 120 TGCAGACATCACTCCAGAGAAATTCACACAGATAGAGACAAAGTGGCCACCCAGCCATCT 61
QY 1784 CTTTCCCTCGGGCTGATTTATCCAGAAATAGGATGTCACCAAGCAACACTTCCAGCC 1843
Db 60 CTTTCCCTCGGGCTGATTTATCCAGAAATAGGATGTCACCAAGCAACACTTCCAGCC 1

RESULT 6
AF478692
LOCUS
DEFINITION
Mus musculus cardiac ankyrin repeat protein (Carp) gene, promoter region and partial cds.
AF478692
AF478692.1 GI:19110906
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 723)
Maeda,T., Sepulveda,J., Chen,H.H. and Stewart,A.F.R.
TITLE
alpha1-Adrenergic activation of the cardiac ankyrin repeat protein gene in cardiac myocytes
Gene 297 (1-2), 1-9 (2002)
2 (bases 1 to 723)
Maeda,T., Sepulveda,J. and Stewart,A.F.R.
Direct Submission
Submitted (29-JAN-2002) Cardiovascular Institute, University of Pittsburgh, 200 Lothrop Street, Pittsburgh, PA 15213, USA
JOURNAL
Location/Qualifiers
1. .723
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="FVB"
/db_xref="taxon:10090"
1. .>723
/gene="Carp"
1. .720
/gene="Carp"
/note="contains promoter and 5' UTR"
<721. .>723
/gene="Carp"
/product="cardiac ankyrin repeat protein"
721. .>723
/gene="Carp"
/codon_start=1
/product="cardiac ankyrin repeat protein"
/protein_id="AAL85342.1"

/db_xref="GI:19110907"
/translation="N"
BASE COUNT 187 a 185 c 180 g 171 t
ORIGIN
Query Match 27.4%; Score 646.2; DB 10; Length 723;
Best Local Similarity 97.5%; Pred. No. 1.5e-177;
Matches 699; Conservative 0; Mismatches 13; Indels 5; Gaps 4;
QY 1601 GAACACGGTGAGCGCTGTGGTGCACCTAAATATGGCCAGTGACACCATAGAGTCAAAAGTGC 1660
Db 1 GAACACGGTGAGCGCTGTGGTGCACCTAAATATGGCCAGTGACACCATAGAGTCAAAAGTGC 60
QY 1661 TTACTGAATGCTTTCAATTTCTCCTAATCTCGTACGATGGCATGTCACAGGGCCATTTT 1720
Db 61 TTACTGAATGCTTTCAATTTCTCCTAATCTCGTACGATGGCATGTCACAGGGCCATTTT 120
QY 1721 AGCTCAGACATCACTCCAGAGAAATTCACACAGATAGAGACAAAGTGGCCACCCAGACCCA 1780
Db 121 AGCTCAGACATCACTCCAGAGAAATTCACACAGATAGAGACAAAGTGGCCACCCAGACCCA 180
QY 1781 TCTCTCTCCCTCGGGCTGATTTATCCAGAAATAGGATGTCACCAAGCAACACTTCCCA 1840
Db 181 TCTCTCTCCCTCGGGCTGATTTATCCAGAAATAGGATGTCACCAAGCAACACTTCCCA 237
QY 1841 GCCAACTGGAGTGTGATAAGTCCAGATTATCAGAAAGATATGGCTGTAAAGTGTGATGCAC 1900
Db 238 GCCAACTGGAGTGTGATAAGTCCAGATTATCAGAAAGATATGGCTGTAAAGTGTGATGCAC 297
QY 1901 AGTGCTTGCAATTTCTTGATACGTTAGTACATATAGAGAGTGCACAAAGAGGAAAAAGAGC 1960
Db 298 AGTGCTTGCAATTTCTTGATACGTTAGTACATATAGAGAGTGCACAAAGAGG-AAAAGAGC 356
QY 1961 AGCGATGTGGTGCATATTAACAGGACGCTGTCCCTGGCTTCCGATACGTTGGATGAC 2020
Db 357 AGCGATGT-GTGCAATATTAACAGGACGCTGTCCCTGGCTTCCGATACGTTGGATGAC 415
QY 2021 TCGCATTTGCTGAGCGGTGTGGTCACTGTCACAAAGGAATGACCTCTCACATTTTCTTCTCGA 2080
Db 416 TCGCATTTGCTGAGCGGTGTGGTCACTGTCACAAAGGAATGACCTCTCACATTTTCTTCTCGA 475
QY 2081 TTCGCATACGCGCGCGCGCAGCTTGTCACTCCCTCTTGGGTTCCACAGACATAAGTCTG 2140
Db 476 TTCGCATACGCGCGCGCGCAGCTTGTCACTCCCTCTTGGGTTCCACAGACATAAGTCTG 535
QY 2141 GAATCAAAATTCACCTGCTCTGAATGGCCACTGGTGGGGCAGGGGTGCTACATTGGCT 2200
Db 536 GAATCAAAATTCACCTGCTCTGAATGGCCACTGGTGGGGCAGGGGTGCTACATTGGCT 595
QY 2201 TCCAGAGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAACGGCTGCTGTGGAGGGG 2260
Db 596 TCCAGAGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAACGGCTGCTGTGGAGGGG 555
QY 2261 CTCACAGGGCCAGTTCCAGGGGTTTCATCCACAGAGAGAGAAAAACATAGACTCGAGG 2317
Db 656 CTCACAGGGCCAGTTCCAGGGGTTTCATCCACAGAGAGAGAAAAACATAGACTCACGG 712

RESULT 7
BV076484
LOCUS
DEFINITION
S212P60369FF7.T0 CZECHII/Ei Mus musculus STS genomic, sequence tagged site.
BV076484
BV076484.1 GI:31192279
ACCESSION
VERSION
KEYWORDS
SOURCE
Mus musculus (house mouse)
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 665)
Wade,C.M., Kulbokas,E.J. III, Kirby,A.W., Zody,M.C., Mullikin,J.C., Lander,E.S., Lindblad-Toh,K. and Daly,M.J.

TITLE The mosaic structure of variation in the laboratory mouse genome
JOURNAL Nature 420 (6915), 574-578 (2002)
MEDLINE 22354684
PUBMED 12466852
COMMENT

Contact: Kerstin Lindblad-Toh
Whitehead Institute for Biomedical Research, Center for Genome
Research
220 Charles Street, Cambridge, MA 02141, USA
Tel: 6172521477
Fax: 6172580903
Email: kersli@genome.wi.mit.edu
Primer A: None
Primer B: None
STS size: 665

Protocol:
WGS-discovery: Paired-end low-coverage whole genome shotgun reads
were generated from 129S1/SvimJ, C3H/HeJ, and BALB/cByJ. The WGS
reads were placed uniquely on the MGSCV3 C57BL/6J assembly and SNP
detection was carried out by SSNA-SNP. 225,000 reads were
annotated
as STSS and 81,000 SNPs were annotated with alleles from C57BL/6J
and the strain from which the particular read came. The validation
rate for these SNPs was estimated at approximately 98%.

FEATURES
source
1..665
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="CZECHII/Ei"
/db_xref="taxon:10090"
/map="19 22-610 35751955-35751368"
/clone.lib="CZECHII/Ei"
<1..>665
191 a 155 c 164 g 155 t

STS
BASE COUNT 191 a 155 c 164 g 155 t
ORIGIN

Query Match 27.2%; Score 640.2; DB 11; Length 665;
Best Local Similarity 98.6%; Pred. No. 8.3e-176;
Matches 656; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 524 GTAGGAACCTGTATGTACCAAGTCTGCTTCTATGGTGGAGCCAGCAGCATCGTGGGTG 583
DB 1 GTAGGAACCTGTATGTACCAAGTCTGCTTCTATGGTGGAGCCAGCAGCATCGTGGGTG 60

QY 584 GAGCGAAGCGCAACTCTACCTCTAGCTCTGCATCCATAGCAAGTAGCCTATGTTTCT 643
DB 61 GAGCGAAGCGCAACTCTACCTCTAGCTCTGCATCCATAGCAAGTAGCCTATGTTTCT 120

QY 644 GTGTCTAGGTGTCTCTGTGAATCGAGATCCTTGGCCTTGTCTTCAATTAGGAGGCAC 703
DB 121 CTGTCTAGGTGTCTCTGTGAATCGAGATCCTTGGCCTTGTCTTCAATTAGGAGGCAC 180

QY 704 AAAATCTCAGAGATTCAAGACTGTCTCAGCAGCCAGAGTCTTCTCTCAAGGAAAGGTC 763
DB 181 AAAATCTCAGAGATTCAAGACTGTCTCAGCAGCCAGAGTCTTCTCTCAAGGAAAGGTC 240

QY 764 TCACTCTCAGCCCCCTTAGCTCTAGCTCAGCCCTGGAAACAAACGGCCACAGGAATGAG 823
DB 241 TCACTCTCAGCCCCCTTAGCTCTAGCTCAGCCCTGGAAACAAACGGCCACAGGAATGAG 300

QY 824 AAAAGCTGCATAGCTGTGTTCACCTTCAAGAGTCAAGAAATAGTGTAAACCATGAA 883
DB 301 AAAAGCTGCATAGCTGTGTTCACCTTCAAGAGTCAAGAAATAGTGTAAACCATGAA 360

QY 884 AACGAGAAGCCAAAGTATTCAATTGATAGCGTCTCAGGACAGATAGGACAGAGAAAC 943
DB 361 AACGAGAAGCCAAAGTATTCAATTGATAGCGTCTCAGGACAGATAGGACAGAGAAAC 420

QY 944 ACTAGAGAGGGGAACCCAGGAAGCAAGGATTATAGTGTGTTTTCAGGGCAATGT 1003
DB 421 ACTAGAGAGGGGAACCCAGGAAGCAAGGATTATAGTGTGTTTTCAGGGCAATGT 480

QY 1004 CTTGTACTCAAGATTCTAGAAACACAATTTGCTGTGTTGAACAGCTCAAGT-GGGGTGGGG 1062

DB 481 CTCGTTCTGAAGGTCTAGAAACACAATTTGCTGGTTGAACGCTGAAGTGGGGTGGG 540
QY 1063 GTTCTTACCCCATGTTTCATGGAAGGTGAGTGAGGAGACAGATATATATGAGCCAGCA 1122
DB 541 GTTCTTACCCCATGTTTCATGGAAGGTGAGTGAGGAGACAGATATATATGAGCCAGCA 600

QY 1123 TAACAACATACACACACACCCTAATTAACACTTCCTCTTCTACTGACACCCCTTCACT 1182
DB 601 TAACAACATACACACACACCCTAATTAACACTTCCTCTTCTACTGACACCCCTTCACT 660

QY 1183 CTCCT 1187
DB 661 CTCCT 665

RESULT 8
AX468604
LOCUS AX468604 2074 bp DNA linear PAT 16-JUL-2002
DEFINITION Sequence 2 from Patent WO0246220.
ACCESSION AX468604
VERSION AX468604.1 GI:21901403
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
AUTHORS Schwartz,B., Branellec,D. and Chien,K.
TITLE Sequences upstream of the carp gene, vectors containing them and
uses thereof
JOURNAL Patent: WO 0246220-A 2 13-JUN-2002;
Aventis Pharma S.A. (FR); The Regents of The University of
California at San Diego (US); Benoit, Patrick (FR)
LOCATION/Qualifiers
source
1..2074
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

BASE COUNT 612 a 469 c 416 g 572 t 5 others
ORIGIN

Query Match 15.6%; Score 367.8; DB 6; Length 2074;
Best Local Similarity 76.4%; Pred. No. 5.9e-96;
Matches 597; Conservative 0; Mismatches 162; Indels 22; Gaps 11;

QY 1540 TCTCTCTGTGCATCACTTGGCCCGTTTGGGGT--AGATCCTCTGATTAGCCTTCAGAT 1597
DB 1289 TCTATTTCTTGACCACTCTGATCCATTTTGAAGTAAATGCTCCAATTATTATGCTGT 1348

QY 1598 TTAGAACACGGTGAGCCTGTGTGTCACCTAATTTATGGCCAGTCACACCATAGAGTCAAAGT 1657
DB 1349 TTAGAACACGGTGAAGCATGTCTATGTGCTA--ATGGCCAGTGACATCATAAAGAAAGT 1405

QY 1658 GCATTACTGAATGCTTTCAATTTCTCTAATCTGTGTAGTGGCATGTCCAGGGCCAT 1717
DB 1406 GCATTACTGAATGCTTTCAATGCTTTAATGATGTTAGTGGCATGTCTAGGGCCCTA 1465

QY 1718 TTTAGCTGCAGACATCACTCCAGAGAAATTCAAAACAGATAGAGACAAAGTGGCACCCAGAC 1777
DB 1466 TTTAGC-CCAGACATCACTCCAAAGAAATTCAAAACAGATATAGACAAGTGCCTTTAGGGC 1524

QY 1778 CCATCTCTCTCCCTCGGGCTGATTTATCCCAAGAAATAGGATGTCCCAAGCAACACTTC 1837
DB 1525 CCAGATCTCTCTCCCTCGGGCTGTTTACCCAGGGAATAGGATGTCTCTGGCAACAGTTTCC 1584

QY 1838 CCAGCAACTGGAGTGTGATAAGTCCAGTTTATCAGAAAGATATGCTCTTAAGTGTGATG 1897
DB 1585 C---CTAAGTGAAGTGTGATAAGTCTGCTTATCAGAAAGATATTACTGGGGGTGTGATA 1641

QY 1898 CACAGTGC--TTGCAATTTTCTTGATACGTTAGCTCATATGAGAGCTGACAAAGAGGAAA 1955
DB 1642 TGTAGGGCATCTACATTTTCTTGATA-GGTAGTCTATGAAAGCTGACAAAGAA--AAAA 1698

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Tracey.A.
Direct Submission
Submitted (19-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquary@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On May 31, 2001 this sequence version replaced gi:14161205.
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP database can be found at
http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human Chromosome 10, constructed by the Sanger Centre Chromosome 10 Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr10
RP11-320F15 is from the library RPC1-11.2 constructed by the group of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBACE3.6
IMPORTANT: This sequence is not the entire insert of clone RP11-320F15 It may be shorter because we sequence overlapping sections only once, except for a short overlap.
The true left end of clone RP11-236B18 is at 50012 in this sequence. The true right end of clone RP11-103A2 is at 100 in this sequence.

Location/Qualifiers
1..50111
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-320F15"
/clone_lib="RPC1-11.2"
complement(1..98)
/note="match: GSS: Em:AQ670367"
complement(1..97)
/note="match: GSS: Em:AQ544636"
complement(1..77)
/note="match: STS: Em:G56439
match: GSS: Em:AQ311708"
120..34169
/gene="BA320F15.1"
join(120..472,3273..3328,3962..4018,4428..4502,7467..7538,14216..14305,23179..23295,23858..23887,24284..24321,24717..24796,28974..29381)
/gene="BA320F15.1"
/product="BA320F15.1.1 (ribonuclease P (30kD) (RPP30))"
/notes="match: CnAAS: Em:U77685 Em:AK004137 Em:U95123
match: ESTs: Em:BE572689 Em:BE378859 Em:BE137893
Em:AA920575 Em:BG106820 Em:BF681617 Em:BF248001
Em:BF031745 Em:AA854455 Em:AI359795"
/evidence="not experimental"
join(391..472,3273..3328,3962..4018,4428..4502,7467..7538,14216..14305,23179..23295,23858..23887,24284..24321,24717..24796,28974..29381)
/gene="BA320F15.1"
/note="match: proteins: Tr:P78346 Tr:O88796"

/codon_start=1
/evidence="not experimental"
/product="BA320F15.1.1 (ribonuclease P (30kD) (RPP30))"
/protein_id="CAC70100.1"
/db_xref="GI:15717970"
/translation="MAVFADLDLRAGSLKALRGVTAHLGYSVVAINHIVDFKEK
KQIEKPVAVSELFTTLIVQKSRPIKILTRLTIIIVSDPSHCNVRATSSRLYDV
VAVRPTEKLFHIACTHLDVLVCITVTEKLPFVKRPINVAIDRGIAFEVLVYPAI
KDSIMRRYTISSALNMOICGKNVLISSAAERPLEIRGPDVANIGLLFGLSESDAK
AAVSNCRALUHGTRTAGLIISTVVKPRESEGDCLPASKAKCEG"
689..800
/note="MIR repeat: matches 30..132 of consensus"
1443..1488
/note="L2 repeat: matches 2459..2502 of consensus"
1761..2155
/gene="BA320F15.1"
/note="match: GSS: Em:AQ805524"
1907..2434
/gene="BA320F15.1"
/note="match: GSS: Em:AQ775672"
1926..2208
/gene="BA320F15.1"
/note="match: GSS: Em:B92129"
2349..2482
/note="L2 repeat: matches 2557..2696 of consensus"
2939..3167
/note="MIR repeat: matches 13..262 of consensus"
complement(4176..4649)
/note="match: GSS: Em:AQ626500"
join(4648..4764,7467..7538,14216..14305,23179..23887,
24284..24321,24717..24796,28974..29381)
/gene="BA320F15.1"
/product="BA320F15.1.2 (putative isoform 2)"
/note="match: ESTs: Em:AW939965"
/evidence="not experimental"
4733..4858
/note="L2 repeat: matches 2601..2732 of consensus"
4928..5102
/note="L2 repeat: matches 2299..2478 of consensus"
5257..5439
/note="L2 repeat: matches 1808..1971 of consensus"
5440..5721
/note="AluX repeat: matches 7..293 of consensus"
5722..6237
/note="L2 repeat: matches 1200..1808 of consensus"
6884..7196
/note="AluDb repeat: matches 1..312 of consensus"
8243..8404
/note="2 copies 81 mer 81% conserved"
8445..8541
/note="HY1 repeat: matches 16..112 of consensus"
8542..8583
/note="HY1 repeat: matches 1..42 of consensus"
8850..9130
/note="AluX repeat: matches 1..289 of consensus"
9601..9741
/note="MIR repeat: matches 76..212 of consensus"
10798..11076
/note="AluY repeat: matches 1..291 of consensus"
11086..11219
/note="FLAM C repeat: matches 1..129 of consensus"
11319..11491
/note="MERSA repeat: matches 3..189 of consensus"
11928..11984
/note="L2 repeat: matches 1743..1801 of consensus"
13755..13817
/note="L2 repeat: matches 2684..2749 of consensus"
15390..15671
/note="AluDb repeat: matches 1..282 of consensus"
15791..16088
/note="AluSg repeat: matches 1..296 of consensus"
16979..17036
/note="L2 repeat: matches 2423..2481 of consensus"

REFERENCE
AUTHORS
TITLE
JOURNAL

COMMENT

FEATURES

source

misc_feature

misc_feature

misc_feature

gene

mRNA

CDS


```
misc_feature 17236..17755
/feature="BA320F15.1"
/feature="match: GSS: Em:AQ551877"
17709..17972
/feature="AluX repeat: matches 45..304 of consensus"
19198..19363
/feature="LINE repeat: matches 5272..5435 of consensus"
19489..19928
/feature="MLTIC repeat: matches 5..478 of consensus"
21838..22258
/feature="L2 repeat: matches 2109..2669 of consensus"
22440..22946
/feature="BA320F15.1"
/feature="match: GSS: Em:AQ876596"
Join(23179..23295,23858..23887,24284..24321,24717..24796,
28974..29072,30647..30748,31596..31680,33706..34469)
/feature="BA320F15.1"
/product="BA320F15.1.3 (putative isoform 3)"
/feature="match: ESTs: Em:BE567341 Em:AI292002 Em:BF684192
Em:BF435407 Em:AI740881 Em:AA651912 Em:BF438915"
/evidence="not experimental"
24121..24518
/feature="BA320F15.1"
/feature="match: GSS: Em:AQ26750"
complement(join(24284..24321,24724..24796,28974..29240))
/feature="match: STS: Em:G24365"
24549..25020
/feature="BA320F15.1"
/feature="match: GSS: Em:AQ223714"
24561..24956
/feature="match: GSS: Em:AQ223714"
25321..25458
/feature="match: GSS: Em:BA53566"
/feature="FLAM_C repeat: matches 1..133 of consensus"
complement(25417..25863)
/feature="match: GSS: Em:AQ712354"
26204..26603
/feature="L1MA2 repeat: matches 5888..6302 of consensus"
28442..28516
/feature="L2 repeat: matches 2636..2710 of consensus"
29149..29154
/feature="BA320F15.1"
29166
/feature="BA320F15.1"

Query Match 14.7%; Score 347.6; DB 9; Length 50111;
Best Local Similarity 78.2%; Pred. No. 78-90;
Matches 521; Conservative 0; Mismatches 129; Indels 16; Gaps 8;

QY 1653 AAAGTGCAATGCTTCAATTTCTCCTAATGCTGGTACGATGGCATGTGCACAGG 1712
DB 50106 AAAGTGCAATGCTTCAATTTCTTATATGATGGTAAGTGCGCATGTCTGGG 50047
QY 1713 GCCATTTTACGTCGACATCTCCAGAGAAATCCAAACAGATAGACAAAGTGGCACC 1772
DB 50046 GCCTATTTAGCCCCACATCACTCCAAAGAAATCCAAACAGATATAGACAAAGTGCCTTT 49987
QY 1773 CAGACCCATCTCTTCCCTCGGCTGATATCCCGAGAAATAGGATGTCCTCCAAAGCAAC 1832
DB 49986 AGGGCCCCAGATCCCTTCCCTCAGGCTGTTTACCAGGGAATAGGATGTCCTTCCGAC 49930
QY 1833 ACTTCCCGACCACTGGAGTGTGATAGTCCTCAGATTTATCAGAAAGATATGGCTGTAAAGTG 1892
DB 49929 AGTTTCCCTTAAGTGAAGTGTGATAGTCTTATCAGAAAGATATTAATCTGGGGGTG 49870
QY 1893 TGATGCACAGTGC--TTGCATTTTCTTGATAGTGTAGTATATAGAGCTGACAAAGAG 1950
DB 49869 TGATATGTAGGCACTACATTTTCTTGATA--GGTAGTTCATATGAAGAGCTGACAAAGAA- 49812
QY 1951 GAAAGAGCAGCGATGGTGCATATTAACAGCAGCTGTCCCTGGCTTCCCGATAC 2010
DB 49811 -AAAAAGGCGAGTGTGGTGCATATGTCAACAGACAGCTGTCCCTGAC-TCTTGGCAA 49754

Query Match 5.5%; Score 128.6; DB 6; Length 1901;
Best Local Similarity 82.6%; Pred. No. 6.9e-26;
Matches 171; Conservative 0; Mismatches 34; Indels 2; Gaps 2;

QY 2105 TCATCTCCCTCTTGGGCTTCCAGACACTAGTCTGGAAATGAAATTCACCTGCCTCTGA 2164
DB 27 TCCCTCCCTCTTACGCTTCCAGACACTGATTCGGAATGAAATTCACCTGCCTCTGA 86
QY 2165 ATTGCCCACTGGTGGGCGAGGGGTGCTGCTTCCAGGCTGGAGATTAATCTAC 2224
DB 87 GTTGGCTCTTAATGGGGTGGAGTGTACTTCGGTTCAGGTTCCAGGTTGGAAGATTAATCTAC 146
QY 2225 CCAGCCCTAGCTATATAA--CGGGCTGGTGTGAGGGGCTTCCAGGGCCAGTTCACGGGG 2283
DB 147 CGGGCCCCAGCTATATAAGCTGACCGGTGTGAGGGGCCAGCAGGGCCCACTCCAGGGA. 206
QY 2284 TTCATCCAAAGAGAGAAAAACATAGA 2310
DB 207 TTCCTTC-CAGCACAGAAAAACATACA 232

RESULT 12
BD094076
LOCUS BD094076 1901 bp DNA linear PAT 27-AUG-2002
DEFINITION Shear stress-responsive DNAs.
```

ACCESSION BD094076
VERSION WD094076.1 GI:22639664
KEYWORDS WO 0125427-A/37
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1901)
AUTHORS Nojima H., Yoshisue H., Obayashi M., Ota T., Kawabata A., Sakurada K., Kuga T., Sekine S., Nakamura Y. and Sugan S.
TITLE Shear stress-responsive DNAs
JOURNAL Patent: WO 0125427-A 37 12-APR-2001;
KYOWA HAKKO KOGYO CO LTD, HIROSHI NOJIMA, HAJIME YOSHISUE, MASAYA OBAYASHI, TOSHIO OTA, AYAKO KAWABATA, KAZUHIRO SAKURADA, TETSURO KUGA, SUSUMU SEKINE, YUSUKE NAKAMURA, SUMIO SUGANO
COMMENT OS Homo sapiens (human)
FN WO 0125427-A/37
PD 12-APR-2001
PF 02-OCT-2000 WO 2000JP006840
PR 01-OCT-1999 JP 99P 280976
PI HIROSHI NOJIMA, HAJIME YOSHISUE, MASAYA OBAYASHI, TOSHIO OTA, PI AYAKO KAWABATA, TETSURO KUGA, SUSUMU SEKINE, YUSUKE NAKAMURA, KAZUHIRO SAKURADA, TETSURO KUGA, SUSUMU SUGANO
PC C12N15/12, C07K14/435, C07K16/18, C12P21/02, C12Q1/68, A61K38/00, A61K39/395, A61K48/00, A61P9/10, G01N33/50, G01N33/53
CC

FEATURES
source
1. .1901
/location/Qualifiers
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 592 a 378 c 460 g 471 t
ORIGIN

Query Match 5.5%; Score 128.6; DB 6; Length 1901;
Best Local Similarity 82.6%; Pred. No. 6.9e-26;
Matches 171; Conservative 0; Mismatches 34; Indels 2; Gaps 2;

QY 2105 TCATCTCCCTCTTGGGCTTCCAGACACACTAAGTCTGGAATGAAATTCACCTGCCTCTGA 2164
DB 27 TCCCTCCCTCTTACGCTTCCAGACACACTGATCTGGAATGAAATTCACCTGCCTCTGA 86

QY 2165 ATTGGCCACTGTGGGGGCGGGGTGATCTGGCTTCCAGCTGGAAGATTATCTCAC 2224
DB 87 GTTGGCTCTTAATGGGGGTGGGAGTGTACTTCGGTTCAGGTTCCAGGTTATCTCAC 146

QY 2225 CCAGCCCTAGCTATATAA-CGGGCTGTGTGGAGGGGCTCCACAGGGCCAGTTCACAGGG 2283
DB 147 CCGGCCCCAGCTATATAAGCTACCGGTGTGGAGGGGCCAGCAGGGCCAACTCCAGGGA 206

QY 2284 TTCATCCACAAGAGAGAAAAACATAGA 2310
DB 207 TTCCTTC-CACGACAGAAAAACATACA 232

RESULT 13
HSRNACINP 1901 bp mRNA linear PRI 05-SEP-1995
LOCUS H.sapiens mRNA for cytokine inducible nuclear protein.
DEFINITION X83703
ACCESSION X83703.1 GI:793840
VERSION ankyrin-like repeat; nuclear localisation signal; nuclear protein.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1
AUTHORS Chu, W., Burns, D.K., Swerlick, R.A. and Presky, D.H.
TITLE Identification and characterization of a novel cytokine-inducible

JOURNAL nuclear protein from human endothelial cells
MEDLINE J. Biol. Chem. 270 (17), 10236-10245 (1995)
PUBMED 95247734
7730328
REFERENCE 2 (bases 1 to 1901)
AUTHORS Chu, W.
TITLE Direct Submission
JOURNAL Submitted (05-JAN-1995) W. Chu, Hoffmann-La Roche, 340 Kingsland Street, Dept. of Inflammation/Autoimmune Disease, Hoffmann-La Roche, Nutley, NJ 07110, USA
FEATURES
source
1. .1901
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="C-193"
/cell_type="endothelial"
/tissue_type="skin"
/clone_lib="HMEC cDNA"
misc_feature 94..98
repeat_unit 152..283
CDS 250..1209
/note="cytokine-inducible expression"
/codon_start=1
/product="nuclear protein"
/protein_id="CAA58676.1"
/db_xref="GI:793841"
/translation="MMVLKVELVTGKKNGCEAGFLPEDPRDGEYEAATLEKQED LKLLAHPVITGEGQWKEKREAEALPKKLEQSKLENLEDELEIIQKKKKYKRT KPVVVEPEPEIITPEVDPTFLKAALENKLEPVKFLSDKNNPDVCDYKRTALHRA CLEGLAIVEKLMEEAGQIEFRDMLLESTAIHWASGGLDLVLLKLNKGAKISARDKL LSTALHVAVRTGHVEGASHLIACEADLNKAREGDTPLHDAVRLNRYKMRILLIMYGA DLNINKCAGKTPMDLVLHWQGTGKAI FDSLRENSYKTSRIATF"
BASE COUNT 592 a 378 c 460 g 471 t
ORIGIN

Query Match 5.5%; Score 128.6; DB 9; Length 1901;
Best Local Similarity 82.6%; Pred. No. 6.9e-26;
Matches 171; Conservative 0; Mismatches 34; Indels 2; Gaps 2;

QY 2105 TCATCTCCCTCTTGGGCTTCCAGACACACTAAGTCTGGAATGAAATTCACCTGCCTCTGA 2164
DB 27 TCCCTCCCTCTTACGCTTCCAGACACACTGATCTGGAATGAAATTCACCTGCCTCTGA 86

QY 2165 ATTGGCCACTGTGGGGGCGGGGTGATCTGGCTTCCAGCTGGAAGATTATCTCAC 2224
DB 87 GTTGGCTCTTAATGGGGGTGGGAGTGTACTTCGGTTCAGGTTCCAGGTTATCTCAC 146

QY 2225 CCAGCCCTAGCTATATAA-CGGGCTGTGTGGAGGGGCTCCACAGGGCCAGTTCACAGGG 2283
DB 147 CCGGCCCCAGCTATATAAGCTACCGGTGTGGAGGGGCCAGCAGGGCCAACTCCAGGGA 206

QY 2284 TTCATCCACAAGAGAGAAAAACATAGA 2310
DB 207 TTCCTTC-CACGACAGAAAAACATACA 232

RESULT 14
G28603
LOCUS human STS SHGC-35401, sequence tagged site.
DEFINITION G28603
ACCESSION G28603
VERSION G28603.1 GI:1408418
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.


```
REFERENCE 1 (bases 1 to 1901)
AUTHORS Myers,R.M.
JOURNAL Unpublished (1996)
COMMENT

Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 4157259687
Fax: 4157259689
Email: myers@shgc.stanford.edu

Primer A: GGCATTTTGAAGGATGG
Primer B: CCAGATGATCATGAAGG
STS size: 222
PCR Profile:
  Initial incubation: 94 degrees C for 90 seconds
  Denaturation: 94 degrees C for 15 seconds
  Annealing: 62 degrees C for 23 seconds
  Polymerization: 72 degrees C for 30 seconds
  PCR Cycles: 30
  Thermal Cycler: Perkin Elmer 9600

Protocol:
  Template: 25 ng
  Primer: each 1 uM
  dNTPs: each 200 uM
  Tag Polymerase: 0.05 units/ul
  Total Vol: 10 ul

Buffer:
  MgCl2: 2.5 mM
  KCl: 50 mM
  Tris-HCl: 20 mM
  pH: 8.3

Prepared with primer pairs provided by Sandoz, derived from X83703
-- Washington University/Merck EST sequence.

FEATURES
    source
        1..1901
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /map="10"
    STS
        1246..1467
    primer_bind
        1246..1263
    primer_bind
        complement(1447..1467)
    BASE COUNT
        592 a 378 c 460 g 471 t
    ORIGIN
        Query Match
        Best Local Similarity 5.5%; Score 128.6; DB 11; Length 1901;
        Matches 171; Conservative 0; Mismatches 34; Indels 2; Gaps 2;
        QY 2105 TCATCTCCCTCTTGGGCTTCCAGACACTAAGCTTGGAAATGAAATTCACCTGCCTCTGA 2164
        |||
        Db 27 TCCCTCCCTCTTACGTTCCAGACACTGATTCTGGAATGAAATTCACCTGCCTCTGA 86
        |||
        QY 2165 ATTGGCCACTGTGGGGCAGGGGTGTGACTTGGCTTCCAGCTGGAAGATTATCTCAC 2224
        |||
        Db 87 GTTGGCTCTTAATGGGGTGGGAGTGTACTTCGGTTCCAGGTTGGAAGATTATCTCAC 146
        |||
        QY 2225 CCAGCCCTAGCTATATAA-CGGGCTGGTGTGAGGGGTCCACAGGGCCAGTTCACGGGG 2283
        |||
        Db 147 CCGGCCCCAGCTATATAGCTACCGGTGTGGAGGGGCCACAGGGCCAACTCCAGGGA 206
        |||
        QY 2284 TTCATCCCAAGAGAAAAACATAGA 2310
        |||
        Db 207 TTCCTTC-CACGACAGAAAAACATACA 232
        |||
    RESULT 15
    AC074094/c
```

LOCUS AC074094
DEFINITION Homo sapiens chromosome 10 clone RP11-236B18, WORKING DRAFT
SEQUENCE, 17 unordered pieces.
ACCESSION AC074094
VERSION AC074094.3 GI:9958197
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
1 (bases 1 to 160350)
Waterston,R.H.
The sequence of Homo sapiens clone
Unpublished
REFERENCE 2 (bases 1 to 160350)
Waterston,R.H.
Direct Submission
Submitted (13-JUL-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
On Sep 1, 2000 this sequence version replaced gi:9665205.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
----- Project Information -----
Center project name: H.NH0236B18
----- Summary Statistics -----
Sequencing vector: M13; 100%
Chemistry: Dye-primer ET; 100% of reads
Chemistry: Dye-terminator Big Dye; 0% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 149545 bases at least Q40
Consensus quality: 152532 bases at least Q30
Consensus quality: 153924 bases at least Q20
Insert size: 151000; agarose-fp
Quality coverage: 4.85 in Q20 bases; agarose-fp
Quality coverage: 4.71 in Q20 bases; sum-of-contigs

* NOTE: This is a 'working draft' sequence. It currently
* consists of 17 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 1 2326: contig of 2326 bp in length
* 2327 2426: gap of unknown length
* 2427 7519: contig of 5093 bp in length
* 7520 7619: gap of unknown length
* 7620 15286: contig of 7667 bp in length
* 15287 15386: gap of unknown length
* 15387 23050: contig of 7664 bp in length
* 23051 23150: gap of unknown length
* 23151 33663: contig of 10513 bp in length
* 33664 33763: gap of unknown length
* 33764 50133: contig of 16370 bp in length
* 50134 50233: gap of unknown length
* 50234 60888: contig of 17855 bp in length
* 60889 68188: gap of unknown length
* 68189 88133: contig of 19945 bp in length
* 88134 88233: gap of unknown length
* 88234 113397: contig of 25164 bp in length
* 113398 113497: gap of unknown length
* 113498 141732: contig of 28235 bp in length
* 141733 141833: gap of unknown length
* 141834 143187: contig of 1355 bp in length
* 143188 143287: gap of unknown length
* 143288 144942: contig of 1655 bp in length

* 144943 145042: gap of unknown length
* 145043 147286: contig of 2244 bp in length
* 147287 147387: gap of unknown length
* 147387 150852: contig of 3466 bp in length
* 150853 150952: gap of unknown length
* 150953 153550: contig of 2598 bp in length
* 153551 153651: gap of unknown length
* 153651 157257: contig of 3607 bp in length
* 157258 157357: gap of unknown length
* 157358 160350: contig of 2993 bp in length.

FEATURES

source

Location/Qualifiers
1. .160350
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="10"
/clone="RP11-236B18"
misc_feature
1. .2326
/note="assembly_name:Contig10"
misc_feature
2427. .7519
/note="assembly_name:Contig11"
misc_feature
7820. .15286
/note="assembly_name:Contig12"
misc_feature
15387. .23050
/note="assembly_name:Contig13"
misc_feature
23151. .33663
/note="assembly_name:Contig14"
misc_feature
33764. .50133
/note="assembly_name:Contig15"
misc_feature
50234. .68088
/note="assembly_name:Contig16"
misc_feature
68189. .88133
/note="assembly_name:Contig17"
misc_feature
88234. .113397
/note="assembly_name:Contig18"
misc_feature
113498. .141732
/note="assembly_name:Contig19"
misc_feature
141833. .143187
/note="assembly_name:Contig3"
misc_feature
143288. .144942
/note="assembly_name:Contig4"
misc_feature
145043. .147286
/note="assembly_name:Contig5"
misc_feature
147387. .150852
/note="assembly_name:Contig6"
misc_feature
150953. .153550
/note="assembly_name:Contig7"
misc_feature
153651. .157257
/note="assembly_name:Contig8"
misc_feature
157358. .160350
/note="assembly_name:Contig9"
BASE COUNT 47390 a 32724 c 33135 g 45498 t 1603 others
ORIGIN

Query Match 5.4%; Score 127.2; DB 2; Length 160350;
Best Local Similarity 52.5%; Pred. No. 2 9e-25;
Matches 171; Conservative 0; Mismatches 153; Indels 2; Gaps 1;
QY 1540 TCTCTGTGCATCACTCGGCCCGTTTGGGGT--AGATCCTCTGATTAGCCTTCAGAT 1597
DB 113707 TCTATTTCTTGACCACTCTGATCCATTTTGAAGTAAATAATGTCCAATTTATTATGCTGTT 113648
QY 1598 TTAGAACGGGTGAGCTGTGGTGCATAATTATGGCCAGTGACACCATAGAGTCAAAGT 1657
DB 113647 TTAGAACGGGTGAAGCATGTCATGTCGTAATTATGGCCAGTGACATCAATAAGAAAGT 113588
QY 1658 GCATTACTGAATGCTTTCAATTCTCCTAATGTTGATGATGGCATGTCACAGGGCCAT 1717
DB 113587 GCATTACTGAATGCTTTCAATTCTTATTAATGATGGTGGCATGTCATGGGGCCTA 113528
QY 1718 TTTAGCTGCAGACATCACTCCAGAGAAATTCAAACAGATAGACAAAGTGGCACCAGAC 1777
DB 113527 TTTAGCCCCAGACATCACTCCAAAGAAATTCNNNNNNNNNNNNNNNNNNNNNNNNNN 113468

QY 1778 CCATCTCTCTCCCTCGGCGCTGATTATCCCGAGAAATAGGATGTCCCAAGCAACACTTC 1837
DB 113467 NNN 113408
QY 1838 CCAGCCAACTGGAGTGTGCTGATAAGTC 1863
DB 113407 NNNNNNNNNNTGTGAGATGTGTATCTC 113382

Search completed: November 18, 2003, 04:21:58
Job time.: 8596.58 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 23:28:30 ; Search.time 618.23 Seconds
(without alignments)
10295.963 Million cell updates/sec

Title: US-10-005-337A-1

Perfect score: 2358

Sequence: 1 ggatccttcattgttaaca.....caggtcgaggccaccatgg 2358

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N Geneseq_19Jun03.*

- 1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
- 2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
- 3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
- 4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
- 5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
- 6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
- 7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
- 8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
- 9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
- 10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
- 11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
- 12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
- 13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
- 14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
- 15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
- 16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
- 17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
- 18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
- 19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
- 20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
- 21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
- 22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
- 23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
- 24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
- 25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2358	100.0	2358	24	ABV73020 Mouse CARP protein
2	1763.4	74.8	2247	21	AA10406 Murine cardiac ank
3	367.8	15.6	2074	24	ABV73021 Human CARP protein
4	130.2	5.5	1988	22	AA193508 Human polynucleoti
5	128.6	5.5	1901	22	AAH02910 Human shear stress
6	128.6	5.5	1901	24	AAH02910 Human shear stress
7	95	4.0	7280	25	ABZ24607 Mouse gastric antr
8	95	4.0	7280	25	AAZ50815 Mouse pre-AMP-18 g

c	9	89.2	3.8	18105	25	ABZ82337	Mouse histamine re
c	10	88.8	3.8	659158	25	ABX16390	Mouse high growth
c	11	88.6	3.8	48974	20	AAZ55300	Mouse presenilin-1
c	12	83.6	3.5	707	24	ABT09109	Phase-1 Rat CT gen
c	13	82	3.5	659158	25	ABX16390	Mouse high growth
c	14	80.4	3.4	10917	24	ABA97708	Wild-type mouse Wo
c	15	78	3.3	892	18	AAV06155	Viral infection ge
c	16	76.8	3.3	249487	24	ABN85733	Mouse genomic regi
c	17	76.6	3.2	5023	24	ABL56461	Nucleotide sequenc
c	18	76.6	3.2	17758	24	ABL56463	Alphab locus of al
c	19	75.8	3.2	335	24	ABS69028	Novel murine polyn
c	20	75.6	3.2	5990	24	ABK62570	Rat sequence diffe
c	21	75.6	3.2	7280	25	ABZ24607	Mouse gastric antr
c	22	75.6	3.2	7280	25	AAZ50815	Mouse pre-AMP-18 g
c	23	75	3.2	49999	20	AAZ23891	Murine LOBO genomi
c	24	75	3.2	49999	20	AAZ23896	Murine LOBO genomi
c	25	74.8	3.2	37940	20	AAZ01026	Partial mouse PGI
c	26	74.6	3.2	10614	16	AAQ89555	Hamster cholestero
c	27	73.4	3.1	4164	19	AAV12216	Mouse retinoid met
c	28	73.4	3.1	4164	24	AAZ24512	Mouse P450RAI geno
c	29	71.6	3.0	171936	24	ABS56565	Human SULF2 genomi
c	30	69.4	2.9	1889	24	AAZ594903	Human DNA sequence
c	31	69.4	2.9	1889	25	ABX77648	Differentially exp
c	32	68.6	2.9	4072	22	AAZ85818	Murine sequence fr
c	33	68	2.9	1094	20	AAZ57479	Rat U3 gene trap d
c	34	68	2.9	16442	18	AAZ83006	Partial mouse WRN
c	35	65.6	2.8	29604	18	AAZ83005	Partial mouse WRN
c	36	65	2.8	6645	17	AAZ65002	Mouse cell cycle r
c	37	63	2.7	874	20	AAZ57448	Rat U3 gene trap d
c	38	63	2.7	874	20	AAZ57470	Rat U3 gene trap d
c	39	63	2.7	3240	22	AAZ05314	Mouse alpha-1,3 ga
c	40	63	2.7	3537	22	AAZ05315	Mouse alpha-1,3 ga
c	41	62	2.6	7208	21	AAA40866	Murine tumour necr
c	42	61.6	2.6	8402	24	ABK87055	Genomic DNA encodi
c	43	60.6	2.6	29392	19	AAV15422	Mouse poly Ig rece
c	44	60.4	2.6	696	24	ABQ96847	Mouse ES cell rela
c	45	59.8	2.5	347	24	ABQ97071	Mouse ES cell rela

ALIGNMENTS

RESULT 1	ABV73020	ABV73020 standard; DNA; 2358 BP.
ID	ABV73020	standard; DNA; 2358 BP.
XX	ABV73020	
AC	ABV73020	
XX	08-JAN-2003	(first entry)
XX	08-JAN-2003	(first entry)
DE	Mouse CARP protein coding sequence upstream DNA fragment.	
XX	Cardiac ankyrin repeat protein; CARP; cardiac; immunosuppressive;	
KW	antifibrotic; Gene therapy; antisense gene therapy; mouse; db.	
XX	Mus musculus.	
OS	Mus musculus.	
XX	WO200246220-A2.	
PN	WO200246220-A2.	
XX	13-JUN-2002.	
PD	13-JUN-2002.	
XX	05-DEC-2001; 2001WO-EP15412.	
PF	05-DEC-2001; 2001WO-EP15412.	
XX	07-DEC-2000; 2000US-251582P.	
PR	07-DEC-2000; 2000US-251582P.	
XX	(AVET) AVENTIS PHARMA SA.	
PA	(REGC) UNIV CALIFORNIA.	
PA	(BENO) BENOIT P.	
XX	Schwartz B, Branellec D, Chien K;	
PI	WPI; 2002-740642/80.	
XX		

1621	GCAC	TATTTATGGCCAGTGACACATAGAGTCAAGTGCAATTACTGAATGCTTTCAAATTT	1681
1681	CTCC	TAAATGCTGTACGATGCGATGTCACAGGGCCATTTTAGCTGCGAGACATCACTCCAG	1740
1681	CTCC	TAAATGCTGTACGATGCGATGTCACAGGGCCATTTTAGCTGCGAGACATCACTCCAG	1740
1741	AGAA	TTCCAAACAGATAGAGNCAAGTGGCACCCAGACCCATCTCCTTCCCTCGGGCTGA	1800
1741	AGAA	TTCCAAACAGATAGAGNCAAGTGGCACCCAGACCCATCTCCTTCCCTCGGGCTGA	1800
1801	TTAT	TCCCGAGAAATAGAGATGTCCCAAAGCAACACTTCCACGCCAACTGGAGTGCATATA	1860
1801	TTAT	TCCCGAGAAATAGAGATGTCCCAAAGCAACACTTCCACGCCAACTGGAGTGCATATA	1860
1861	GTCC	AGTTATCAGAAAGATATGGCTGTAAAGTGTGATGCACAGTGCCTTTCGCAATTTCTTGAT	1920
1861	GTCC	AGTTATCAGAAAGATATGGCTGTAAAGTGTGATGCACAGTGCCTTTCGCAATTTCTTGAT	1920
1921	ACGT	TAGTCATATGAGAGCTGCACAAAGAGGAAAAAGAGCAGCATGTGTTGCAATATTA	1980
1921	ACGT	TAGTCATATGAGAGCTGCACAAAGAGGAAAAAGAGCAGCATGTGTTGCAATATTA	1980
1981	ACAG	GAGCTGTCCCTGGCTTCCCGATACGTGGGATGACTCCGATTCGTGACGGGTGC	2040
1981	ACAG	GAGCTGTCCCTGGCTTCCCGATACGTGGGATGACTCCGATTCGTGACGGGTGC	2040
2041	GTCA	TGCCAAAAGGAATGACCCCTCTCACATTTCTTCTCGATTTCGATACGCCGGGCCAG	2100
2041	GTCA	TGCCAAAAGGAATGACCCCTCTCACATTTCTTCTCGATTTCGATACGCCGGGCCAG	2100
2101	CTTG	TCATCTCCCTTTGGGCTTCCAGACACTAAAGTCTGGAATGAAAAATTCACCTGCCT	2160
2101	CTTG	TCATCTCCCTTTGGGCTTCCAGACACTAAAGTCTGGAATGAAAAATTCACCTGCCT	2160
2161	CTGA	ATTTGGCCACTGTGTGGGGCAGGGGTGTACTTGGCTTCCAGGCTGGAAGATTATC	2220
2161	CTGA	ATTTGGCCACTGTGTGGGGCAGGGGTGTACTTGGCTTCCAGGCTGGAAGATTATC	2220
2221	TCAC	CAGCCCTAGCTATATAAGGGCTGTGTGGAGGGGCTCCACAGGCCAGTTCCAG	2280
2221	TCAC	CAGCCCTAGCTATATAAGGGCTGTGTGGAGGGGCTCCACAGGCCAGTTCCAG	2280
2281	GGGT	TTATCCCAAGAGAGAAAAACATAGACTTCGAGTCTAGGAGCTTTCGATGCCCTGCA	2340
2281	GGGT	TTATCCCAAGAGAGAAAAACATAGACTTCGAGTCTAGGAGCTTTCGATGCCCTGCA	2340
2341	GGTC	GAGGCCACCATGG	2358
2341	GGTC	GAGGCCACCATGG	2358

RESULT 2

RESUL 2
AAA10406
ID AAA10406 standard; DNA: 2247 BP.

XX
AC AAA10406;

XX
DT 18-JUL-2000 (first entry)

DE Murine cardiac ankyrin repeat protein (CARP) promoter.

XX Cardiac ankyrin repeat protein; CARP promoter; murine; adenovirus vector;
KW cardiac specific; heart disease; gene therapy; ds.

XX
OS Mus musculus.

XX
PN WO200015821-A1.

XX
PD 23-MAR-2000.

XX
PF 10-SEP-1999;

11-SEP-1998; 98US-00999960.
PR
XX

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

(REGC) UNIV CALIFORNIA.

Chien KR, Wang Y, Evans S;

WFI; 2000-271457/23.

Human type-5 recombinant adenovirus vector used for targeted gene PT therapy for heart disease and evaluating gene function contains a PT tissue-restricted promoter and inverted terminal repeat sequences - XX

Claim 8; Page 29-30; 33pp; English.

The invention relates to a human type-5 recombinant adenovirus vector XX for achieving cardiac-restricted transcription of a gene of interest. CC

The vector comprises inverted terminal repeat (ITR) sequences from human CC adeno-associated virus (AAV) type 2 (AAAI0404-AI0405) and a cardiac CC tissue- specific promoter. In particular, the promoter is that of the CC cardiomyocyte- restricted cardiac ankyrin repeat protein (CARP) gene. CC

The adenovirus vector is used for targeted gene therapy for heart CC disease and for evaluating gene function. Cardiac restricted CC transcription of a transgene in both neonatal and mature cardiac tissues CC can be achieved to treat inherited and acquired heart diseases. The CC vector is suitable for tissue-specific use in vivo and in vitro and CC provides cardiac restricted transcription. The present sequence CC represents the murine cardiac ankyrin repeat protein (CARP) promoter. CC

Sequence 2247 BP; 609 A; 549 C; 485 G; 589 T; 15 other; SQ

Query Match 74.8%; Score 1763.4; DB 21; Length 2247;
Best Local Similarity 91.2%; Pred. No. 0;
Matches 2124; Conservative 0; Mismatches 69; Indels 137;

1	Qy	GGATCTCTTTTCATGTTTAAACAATATCAACCCCTAAACCAAGGGGAAACAGCCTCCCTGCAGAGT	60
31	Db	GGATCCTTTTCATGTTTAAACAATATCAACCCCTAAACCAAGGGGAAACAGCCTCCCTGCAGAGT	90
61	Qy	GGCTTTGGCCACCACCATGAATACTTCTAGTCTAGTTCGGTTCTGAAACTCAGCCCATCCCA	120
91	Db	GGCTTTGCACCCCATGAATACTTCTAGTCTAGTTCGGTTCTGAAACTCAGCCCATCCCA	150
121	Qy	ACACTTCTGCAAGCCCCATCTCTACAAGGTGCTCATTTGGAAATTTCTTGAGAGTTCTCT	180
151	Db	ACACTTCTGCAAGCCCCATCTCTACAAGGTGCTCATTTGGAAATTTCTTGAGAGTTCTCT	210
181	Qy	TTTCAGGATCAGGCTGATTTCTAGGCAGCAGATTCTCAACCTGGGGGCTCGACCCCTTTGG	240
211	Db	TTTCAGGATCAGGCTGATTTCTAGGCAGCAGATTCTCAACCTGGGGGCTCGACCCCTTTGG	270
241	Qy	GGGAATCAAAACGACCTTTTACAGGGGTCAACATATCATCTATCTATATGTCAGGATATTTA	300
271	Db	GGGAATCAAAACGACCTTTTACAGGGGTCAACATATCATCTATCTATATGTCAGGATATTTA	330
301	Qy	CATTACGATTCGTAAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTAT	360
331	Db	CATTACGATTCGTAAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTAT	390
361	Qy	GATTGAAGTTCACCAACAACATGAGGCCGCCACACTGTTCTAGAGAAAATACACCTGGGTG	420
391	Db	GATTGAAGTTCACCAACAACATGAGGCCGCCACACTGTTCTAGAGAAAATACACCTGGGTG	450
421	Qy	GGGAAGGTTTGGGAAGCCTTTCTGTCATTCTTCTATTCTTCAAAGTGATGTCGTCACA	480
451	Db	GGGAAGGTTTGGGAAGCCTTTCTGTCATTCTTCTATTCTTCAAAGTGATGTCGTCACA	510
481	Qy	GAAGCCCTTTACGCTGTTTCTGCTGGGCTCTTAGTAAGTCTGAGTAGGAACTGTATGAC	540
511	Db	GAAGCCCTTTACGCTGTTTCTGCTGGGCTCTTAGTAAGTCTGAGTAGGAACTGTATGAC	570
541	Qy	CAGGCTTCGCTTCTTATGGGTGGGCAAGACGCATCGTGGGTGAGCGAGACCGAACCT	600
571	Db	CAGGCTTCGCTTCTTATGGGTGGGCAAGACGCATCGTGGGTGAGCGAGACCGAACCT	630

Qy	601	CACCTTCTAGCTCTGCAGTCCCATAGCAAGTAGCCCTAAATGTTTCTGTCTGTCTAGGTGTCACTCT	660
Db	631	CACCTTCTA-CTCTGCATCCATAGCAAGTAGCCCTAAATG-TTCTGNGTCTAGG-GTCACTCT	687
Qy	661	CTGTGAATCTAGATCCCTTGG-CTTGTGCTTGAATTTAGGAGGCACAAAATACTCAGAGATT	719
Db	688	CTGTGAATCTAGATCTCTTGGCCCTTGTGAATTTAGGAGGCACAAAATCTTAAAAAATT	747
Qy	720	CAAGACTGCTCAGCAGCCACAGA--GTCTTCTCCTCAAAAGGAAGGTCTCAACTCTCAGCCC	777
Db	748	CAAGACTGNTCAACAAANCANNAAGTCTTCTCAAAAGGAAGGNCCTTAACNTWNAACC	807
Qy	778	CCCTTTAGCTCT--GAGTCAGCCCTGGAAC--AAACGGCCACAGGAATGAGAAAAGC----	829
Db	808	CCCTTTACTTTTGAGTCAAGSCCTGGAACAAACCCGCCCCAGGAATGAAAAAGCTTGC	867
Qy	830	--TGCCATAGTGTCTTGTCACTTCAAGAGTCAAAAGAAAATAGTGTTTAACCATGAAAACG	887
Db	868	CATNACCTGTTGGCCCTTTNAANAGNCAAAAAAAAATTTGTGGTTTAAACNTTTGAAAAAC	927
Qy	888	AGAAGACCAACAGTTATCTCATTGTATAGCGTCTCAGGACAGATAGGACAGAGAGAACACTA	947
Db	928	CGAAGACCAACAGTTATCC-----	946
Qy	948	GGAGAGGGGAACCCACGAAGGACAAGGTATTAGTGTGTGTTTTTCAGGGCAATGCTTGT	1007
Db	947	-----	946
Qy	1008	TACTGAAGATTCTAGAAAACACAATTTCTGTTGAAACAGCTGAAGTGGGTGGGGTTCCT	1067
Db	947	-----TCTAGAAAACAATTTCTGTTGAAACAGCTGAAGTGGGTGGGGTTCCT	996
Qy	1068	TACCCCATGTTTCATGGAAGGTGAGTCAGGAGAGACAGATATATGATGGCCAGCATAACA	1127
Db	997	TACCCCATGTTTCATGGAAGGTGAGTCAGGAGAGACAGATATATGA--GGCCAGCATAAACA	1055
Qy	1128	AACATACACAACACCCCTAAATTAACACTTCCCTCTTCTTACTGACACCCCTTTCACCTCCT	1187
Db	1056	AACATACACAACACCCCTAAATTAACACTTCCCTCTTCTTACTGACACCCCTTTCACCTCCT	1115
Qy	1188	CTTTCATAAAAAATAAAAAAGTATTTATATGTTGGCTCTTACGATAGAAATCTTTCCTCGAA	1247
Db	1116	CTTTCATAAAAAATAAAAAAGTATTTTA--GTGGCTCTTACGATAGAAATCTTTCCTCGAA	1174
Qy	1248	CTATAAAAGACTATAATATTTATATTTTACACATTTTAAATATCTTAGCGATGACAAGCC	1307
Db	1175	CTATAAAAGACTATAATATTTATTTTTCACATTTTAAATCTTAGCGATGACAAGCC	1234
Qy	1308	AGAAACAAGTATTTTTGGCTCTCTCAACAGCAAGCTTGGGGCTTTTTTGTTCCTGT	1367
Db	1235	AGAAACAAG-ATTTTTTGGCTCTCTCAACAGCAAGCTTGGGGCTTTTTTGTTCCTGT	1293
Qy	1368	TAGGAATAGAAACAGCAGAGCCCGTGTAATCTAGCAGATGCTTATCATTTAGCCCCATGAG	1427
Db	1294	TAGGAATAGAAACAGCAGAGCCCGTGTAATCTAGCAGATGCTTATCATTTAGCCCCATGAG	1353
Qy	1428	TCTCCAGCCTCAGACGCACATTTTCTCGGGCTCTTCTAGCTTTTCCACAGCATGGG	1487
Db	1354	TCTCCAGCCTCAGACGCACATTTTCTCGGGCTCTTCTAGCTTTTCCACAGCATGGG	1413
Qy	1488	AAACTTTACTGCAGAGCATCCAAGTTGTGCTTCTGCTAAGAACTGGACTCACATCTCTCTG	1547
Db	1414	AAACTTTACTGCAGCATCCAAGTTGTGCTTCTGCTAAGAACTGGACTCACATCTCTCTG	1473
Qy	1548	TGATCATCTTCGGGCCCTTTTGGGGTATAGTCCTCTGATTTAGCCTTACAGATTTAGAACAG	1607
Db	1474	-GCATCATCTTCGGGCCCTTTTGGGGTATAGTCCTCTGATTTAGCCTTACAGATTTAGAACAG	1532
Qy	1608	GTGAGCCTGTGGTGCACCTAATATGGCCAGTGACACCATAGGTCAAGTGCATTTACTGA	1667
Db	1533	GTGAGCCTGTGGT--CACTAATATTGGCCAGTGACACCATAGGTCAAGTGCATTTACTGA	1591
Qy	1668	ATGCTTTCAATTTCTCTCAATATGTTGGTACGATGGCATGTGCACAGGGCCATTTTAGCTGCA	1727

1592	Db		ATCGTTTCAATTTCTCCTAATGCTGTGTAGATGGCATGTACAGGGCCATTTTAGTGCA	1551
1728	Qy		GACATCACTCCAGAGAATTCCAAACAGATAGACACAAGTGGCCACCCAGACCCCATCTCCCTT	1787
1652	Db		GACATCA-TCCAGAGAATTCCAAACAGATAG-GACAAGTGGCACCCAGACCCCATCTCCCTT	1709
1788	Qy		CCCGTGGGCTGATTTATCCCCAGAAATAGAGATGTCCTCCAAAGCAACACTTCCCAAGCCAACT	1847
1710	Db		CCCGTGGGCTGATTTATCCCA-AAATAGAGATGTCCTCCAAAGCAACACTTCCCAAGCCAACT	1768
1848	Qy		GGAGTGCTGATAAGTCCAGTTATCAGAAAGATATGGCTGTAAAGTGTGATGCACAGATGCTT	1907
1769	Db		GGAGTGCTGATAAGTCCAGTTATCAGAAAGATATGGCTGTAAAGTGTGATGCACAGATGCTT	1828
1908	Qy		GCAATTTCTTGATACGTTAGTCATATAGAGAGCTGCACAAAGAGGAGAAAGAGCAGCGATG	1967
1829	Db		GCAATTTCTTGATACGTTAGTCATATAGAGAGCTGCACAAAGAGGAGAAAGAGCAGCGATG	1888
1968	Qy		TGTTGCAATATTAAACAGGCAGCTGCTCCCTGGCTTCCCAGATAGTGGGATGACTTCGCATT	2027
1889	Db		T-GTGCNATATTAACAGGCAGCTGTCCCCTGGCTTCCCAGATAGTGGGATGACTTCGCATT	1947
2028	Qy		GCTGAGCGGTGGTGCTCACTGCCAAAGAAATGACCCCTCTCACATTTCTTCTGTAATTCGCAT	2087
1948	Db		GCTGAGCGGTGGTGCTCACTGCCAAAGAAATGACCCCTCTCACATTTCTTCTGTAATTCGCAT	2007
2088	Qy		ACGCCGGGCACGTTGTCACTCTTCTGGGCTTCCCAGACACTAAAGTCTCGGAATGAA	2147
2008	Db		ACGCCGGGCACGTTGTCACTCTCTTGGGCTTCCCAGACACTAAAGTCTCGGAATGAA	2067
2148	Qy		AATTCACTGCCTCTGAAATTTGGCCACTGTGTGGGGCAGGGGTTGTGACTTGGCTTCCGAGG	2207
2068	Db		AATTCACTGCCTCTGAAATTTGGCCACTGTGTGGGGCAGGGGTTGTGACTTGGCTTCCGAGG	2127
2208	Qy		CTGGAGATTTATCTCACCCAGCCCTAGCTATATAACGGGCTGTGTGGAGGGGCTCCACA	2267
2128	Db		CTGGAGATTTATCTCACCCAGCCCTA-CTATATAACGGGCTGTGTGGAGGGGCTCCACA	2186
2268	Qy		GGGCCAGTTCCAGGGGTTTCATCCAAAGAGAGAAAAACATAGACTCGAGG	2317
2187	Db		GGGCCAGTTCCAGGGGTTTCATCCAAAGAGAGAAAAACATAGACTCACGG	2236

RESULT 3
ABV73021

ABV73021
ID ABV73021 standard: DNA: 2074 BP.

AC ABV73021;

DT 08-JAN-2003 (first entry)

XX
DE Human CARP protein coding sequence upstream DNA fragment.

Cardiac ankyrin repeat protein; CARP; cardiant; immunosuppressive;
XX
KW
antiinflammatory; gene therapy; antisense gene therapy; human; ds.
KW

xx
OS Homo sapiens.

XX PN WO200246220-A2.

XX
PD
73-TTN-2002XX
PF 05-DEC-2001: 2001WO-EP15412XX
07-DEC-2000: 2000115-251582P

XX PA (AVET) AVENTTS PHARMA SA

PA (AVET) } AVENTIS PHARMA S
PA (REGC) } UNIV CALIFORNIA

PA (REGC) UNIV CALI
PA (BENO/) BENOIT P.

XX
PI Schwartz B, Branellec D, Chien K;XX
T7

DR WPI; 2002-740642/80.

PT New promoter sequence derived from a portion upstream of the coding

PT sequence of a gene for Cardiac Ankyrin Repeat Protein, for controlling

PT the level and specificity of expression of a transgene in cardiac

PT muscle cells -

XX Claim 5; Fig 2; 48pp; English.

XX The invention relates to a polynucleotide (I) comprising a fragment of a

CC sequence upstream of the coding part of the gene for the Cardiac Ankyrin

CC Repeat Protein (CARP). (I) is capable of inducing a specific expression

CC in vivo of a gene operably linked to (I), in cardiac cells. (I) or a

CC vector (Iib) comprising (I) is useful for the manufacture of a medicament

CC intended for the treatment of cardiac insufficiency, cardiac hypertrophy

CC and hypoxia, and for preventing rejection during cardiac transplant. An

CC expression cassette under the control of (I) is useful for encoding a

CC protein or RNA which is capable of activating the growth of cardiac

CC cells, reducing or suppressing an immune response, inducing angiogenesis,

CC correcting muscle contractility, cardiac hypertrophy, cardiac

CC insufficiency and myocarditis. (Iib) is useful for expressing a gene of

CC therapeutic interest in vivo, by isolating (Iib) and introducing (Iib) in

CC the cardiac tissue, under conditions so that the gene of interest is

CC expressed. (I), the vectors and the compositions are useful in clinical,

CC experimental, therapeutic and diagnostic fields, and in the treatment and

CC prevention of cardiac pathologies. (I) is also useful for generating

CC transgenic animals which constitute models for studying certain cardiac

CC pathologies. The transgenic animals are also useful for screening

CC molecules for their activity on the regulatory sequences of the gene

CC encoding the CARP protein. The present sequence represents the DNA

CC fragment upstream of the coding sequence of a human CARP protein.

XX

SQ Sequence 2074 BP; 612 A; 469 C; 416 G; 572 T; 5 other;

Query Match 15.6%; Score 367.8; DB 24; Length 2074;

Best Local Similarity 76.4%; Pred. No. 5.9e-97;

Matches 597; Conservative 0; Mismatches 162; Indels 22; Gaps 11;

QY 1540 TCTCTGTGTCATCACTCGGCGCGTTTGGGGT--AGATCCTCTGATTAGCTTCAGAT 1597

DB 1289 TCTATTCTTGACCACTCTGATCCATTTTGAAGTAAATATGTCCTCAATTTATGCTGTT 1348

QY 1598 TTAGAACCGGTGAGCCTGTGTGCTAATATATGGCCAGTGACACCATAGAGTCAAAGT 1657

DB 1349 TTAGAACCGGTAAAGCATGTCTGTCTA---ATGGCCAGTGACATCAATAAGAAAGT 1405

QY 1658 GCATTACTCAATGCTTTTCAATTTCTCTAATGCTGTGATGATGATGTCAGGGCCAT 1717

DB 1406 GCATTACTCAATGCTTTTCAATGCTTTAATGATGATGATGATGATGATGATGATGAT 1465

QY 1718 TTTAGCTGCAGACATCACTCCAGAGAAATCCAAACAGATAGACAAAGTGGCCACCCAGAC 1777

DB 1466 TTTAGC-CCAGACATCACTCCAAAGAAATTTCCAAACAGATATAGACAAAGTCTTTAGGC 1524

QY 1778 CCATCTCTCTCCCTCGGCTGATTATCCCAAGAAATAGATGTCCTCAAGCAACACTTC 1837

DB 1525 CCAGATCCCTTCCCTCAGGCTGTTTACCAGGGAATAGATGTCCTCGGCAAGTTTCC 1584

QY 1838 CCAGCAACTGGAGTCTGATAGTCCAGTTCATATCAGAAAGATATGCTGTAAGTGTGATG 1897

DB 1585 C---CTAAGTGAAGTGTGATAGTCTGCTTATCAGAAAGATATTTACTGGGGGTGTGATA 1641

QY 1898 CACAGTGC--TTGCATTTCTTGATACGTTAGTCATATCAGAGCTGACAAAGAGAA 1955

DB 1642 TGTAGGCACTACATTTTCTTGATA-GGTAGTCATATGAAGCTGACAAAGAA--AAAA 1698

QY 1956 AGACGAGCGATGTGGTCAATATTAACAGCAGCTGTCTCCCTGGCTTCCCGATACGTGGG 2015

DB 1699 AGGCGAGTGTGGTGAATGTCAACAGACAGCTGTCTCCCTGAC-TCITGACAAATAG 1757

QY 2016 ATGACTCGCATGCTGAGCGGTGTGTCTACTGCGCAAGAAAGTACCTCTCAATTTCT 2075

DB 1758 ATGACTTGCATTGCTGAGCGATGTGATCATCAACCAAGAAAGTGGCCCTCTCAATTTCT 1817

2076 CCTGATTGTCATACGCGCGG-----CCAGCTTGTTCATCTCCCTCTTGGGCTTCCAGAC 2130

1818 CCTGATTACATATTCAGCAGGTTAGCTTGTCTCCCTCTTCCCTCTTCCAGCTTCCAGAC 1877

2131 ACTAAGTCTGGAATGAATAATTCACCTGCTCTGAAATTTGGCCACTGTTGGGGGAGGGTG 2190

1878 ACTGAGTCTGGAATGAATAATTCACCTGCTCTGAGTTGGCTCTAATGATGGGGGGAGTG 1937

2191 TCAGTTGGCTTCCAGGCTGGAAGATTATCTCACCCAGCCCTAGCTATATAA-CGGGCTG 2249

1938 TTACTTGGTTCCTCCAGTTTGAAGATTATCTCACCCGCGCCAGCTATATAAGCTGACCG 1997

2250 GTGTGAGGGGCTCCACAGGCGCAGTTCCAGGGGTTTCATCCACAGAGAGAAAAACATAG 2309

1998 GTGTGAGGGGCGCCAGCAGGCGCAACTCCAGGATTCCTTC-CACGACAGAAAAACATAC 2056

2310 A 2310

2057 A 2057

RESULT 4.

AAI93508

ID AAI93508 standard; cDNA; 1988 BP.

XX

AC AAI93508;

XX

DT 06-NOV-2001 (first entry)

XX

DE Human polynucleotide SEQ ID NO 13568.

XX

KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;

KW tissue growth factor; immunomodulatory; cancer; leukaemia;

KW nervous system disorders; arthritis; inflammation; ss.

XX

OS Homo sapiens.

XX

PN WO200164835-A2.

XX

PD 07-SEP-2001.

XX

PF 26-FEB-2001; 2001WO-US04927.

XX

PR 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Liu C, Drmanac RT;

XX

XX WPI; 2001-514838/56.

DR P-PSDB; AAO13577.

XX

PT Isolated nucleic acids and polypeptides, useful for preventing

PT diagnosing and treating e.g. leukaemia, inflammation and immune

PT disorders -

XX

PS Claim 1; SEQ ID NO 13568; 1399pp + Sequence Listing; English.

XX

CC The invention relates to human polynucleotides (AAI79941-AAI93841) and

CC the encoded proteins (AAO00010-AAO13910) that exhibit activity relating to

CC cytokine, cell proliferation or cell differentiation or which may induce

CC production of other cytokines in other cell populations. The

CC polynucleotides and polypeptides are useful in gene therapy, vaccines or

CC peptide therapy. The polypeptides have various cytokine-like activities,

CC e.g. stem cell growth factor activity, haematopoiesis regulating

CC activity, tissue growth factor activity, immunomodulatory activity and

CC activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and

CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 1988 BP; 632 A; 384 C; 472 G; 500 T; 0 other;

Query Match 5.5%; Score 130.2; DB 22; Length 1988;
Best Local Similarity 83.1%; Pred. No. 3.6e-27;
Matches 172; Conservative 0; Mismatches 33; Indels 2; Gaps 2;

QY 2105 TCATCTCCCTCTTGGGCTTCCAGACACTAAAGTCTGGAATGAAAATTCACCTGCCTCTGA 2164
DB 27 TCCCTCCCTCTTTCAGCTTCCAGACACTGAATTCGGAATGAAAATTCACCTGCCTCTGA 86
QY 2165 ATTGGCCACTGTGGGGGCGAGGGGTGACTTGGCTTCCAGGCTGGAAGATTATCTCAC 2224
DB 87 GTTGGCTCTTAATGGGGGGGAGTGTACTTCGGTTCCAGGTTGGAAGATTATCTCAC 146
QY 2225 CCAGCCCTAGCTATATAA-CGGGCTGGTGTGGAGGGGCTCCAGAGGCCAGTTTCCAGGGG 2283
DB 147 CCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGGGGCCAGCAGGGCCCAACTCCAGGGA 206
QY 2284 TTCATCCACAAGAGAGAAAAACATAGA 2310
DB 207 TTCCTTC-CACGACAGAAAAACATACA 232

RESULT 5
AAH02910
ID AAH02910 standard; DNA; 1901 BP.
XX
XX AAH02910;
AC
XX
DT 15-JUN-2001 (first entry)
XX
DE Human shear stress-response coding sequence SEQ ID NO: 73.
XX
KW Human; shear stress-response protein; vascular disease;
KW arteriosclerosis; ds.
XX
XX Homo sapiens.
XX
XX WO200125427-A1.
PN
XX
PD 12-APR-2001.
XX
PF 02-OCT-2000; 2000WO-JP06840.
XX
XX 01-OCT-1999; 99JP-0280976.
PR
XX (KYOW) KYOWA HAKKO KOGYO KK.
PA (NOJI/) NOJIMA H.
XX
XX Nojima H, Yoshisue H, Obayashi M, Ota T, Kawabata A, Sakurada K;
PI Kuga T, Sekine S, Nakamura Y, Sugano S;
PI
XX WPI; 2001-266308/27.
DR P-PSDB; AAB50787.
XX
XX DNA sequences, proteins encoded by them and antibodies against them
PT useful in diagnosis and treatment of vascular disease caused by
PT arteriosclerosis -
XX
XX
PS Claim 20; Page 422-425; 678pp; Japanese.
XX
CC The present invention provides the protein and coding sequences of a
CC number of human shear stress response proteins. These are useful in the
CC diagnosis, treatment and screening of vascular diseases caused by
CC arteriosclerosis, including heart failure, post-PTCA restenosis and
CC hypertension.
XX
SQ Sequence 1901 BP; 592 A; 378 C; 460 G; 471 T; 0 other;

Query Match 5.5%; Score 128.6; DB 22; Length 1901;

Best Local Similarity 82.6%; Pred. No. 1e-26;
Matches 171; Conservative 0; Mismatches 34; Indels 2; Gaps 2;

QY 2105 TCATCTCCCTCTTGGGCTTCCAGACACTAAAGTCTGGAATGAAAATTCACCTGCCTCTGA 2164
DB 27 TCCCTCCCTCTTTCAGCTTCCAGACACTGAATTCGGAATGAAAATTCACCTGCCTCTGA 86
QY 2165 ATTGGCCACTGTGGGGGCGAGGGGTGACTTGGCTTCCAGGCTGGAAGATTATCTCAC 2224
DB 87 GTTGGCTCTTAATGGGGGGGAGTGTACTTCGGTTCCAGGTTGGAAGATTATCTCAC 146
QY 2225 CCAGCCCTAGCTATATAA-CGGGCTGGTGTGGAGGGGCTCCAGAGGCCAGTTTCCAGGGG 2283
DB 147 CCGGCCCCAGCTATATAAGCTGACCGGTGTGGAGGGGCCAGCAGGGCCCAACTCCAGGGA 206
QY 2284 TTCATCCACAAGAGAGAAAAACATAGA 2310
DB 207 TTCCTTC-CACGACAGAAAAACATACA 232

RESULT 6
AAD27217
ID AAD27217 standard; DNA; 1901 BP.
XX
XX AAD27217;
AC
XX
DT 09-APR-2002 (first entry)
XX
DE Human CAA58676 protein encoding EST clone X83703 DNA.
XX
KW Human; congestive heart failure; dilative cardiomyopathy; sudden death;
KW hypertrophic cardiomyopathy; ischaemic cardiomyopathy; rhythm disorder;
KW heart muscle disease; conduction disorder; coronary heart disease;
KW systemic arterial hypertension; pulmonary hypertension; endocarditis;
KW pulmonary heart disease; valvular heart disease; pericardial disease;
KW congenital heart disease; gene therapy; syncope; transgenic animal;
KW expressed sequence tag; EST; clone X83703; CAA58676 protein; ds.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH 250..1209
CDS /tag= a
FT /product= "Human CAA58676 protein"
FT 198..431
FT /tag= b
FT /note= "66268 cDNA fragment"
FT 1604..1753
FT /tag= b
FT /note= "S1MC01-1 cDNA fragment"
FT 1302..1306
FT /tag= b
FT /note= "AU-rich mRNA decay element"
FT 1391..1401
FT /tag= b
FT /note= "AU-rich mRNA decay element"
FT 1415..1423
FT /tag= b
FT /note= "AU-rich mRNA decay element"
XX
XX WO200192567-A2.
PN
XX
XX 06-DEC-2001.
PD
XX
XX 30-MAY-2001; 2001WO-EP06165.
PF
XX
XX 30-MAY-2000; 2000US-207400P.
PR
XX (MEDI-) MEDIGENE.AG.
PA
XX
XX Bunk D, Reuner B, Beck J, Henkel T;
PI
XX WPI; 2002-122073/16.
DR


```
DR P-PSDB; AAE16633.
XX
XX Identifying a subject at risk for a heart disease e.g. congestive heart
PT failure, dilative cardiomyopathy, heart muscle disease, by quantifying
PT the polypeptide expressed by genes abnormally expressed in heart tissue
PT
XX
PS Claim 2a; Fig 10b; 154pp; English.
XX
XX The patent discloses novel target genes abnormally expressed in heart
CC tissues and their corresponding proteins. The invention also relates to
CC methods for assessing the expression level of these genes. The method
CC is used for testing the predisposition of mammals and preferably humans
CC for a heart disease or for an acute state of such a disease. It is also
CC useful to treat diseases of the heart such as congestive heart failure,
CC dilative cardiomyopathy, hypertrophic cardiomyopathy, ischaemic cardio-
CC myopathy, specific heart muscle disease, rhythm and conduction disorders,
CC syncope and sudden death, coronary heart disease, systemic arterial
CC hypertension, pulmonary hypertension, pulmonary heart disease, valvular
CC heart disease, congenital heart disease, pericardial disease and
CC endocarditis. Sequences of the invention are also used in gene therapy.
CC A transgenic non-human mammal comprising the sequences of the invention
CC are useful for the development for medicaments for the treatments of
CC heart diseases. The present DNA sequence is expressed sequence tag
CC (EST) clone X83703 which encodes CAA58676 protein.
XX
SQ Sequence 1901 BP; 592 A; 378 C; 460 G; 471 T; 0 other;
Query Match 5.5%; Score 128.6; DB 24; Length 1901;
Best Local Similarity 82.6%; Pred. NO. 1e-26;
Matches 171; Conservative 0; Mismatches 34; Indels 2; Gaps 2;
QY 2105 TCATCTCCCTCTTGGGCTTCCAGACACTAAAGTCTGGAATGAAATTCACCTGCCTCTCA 2164
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2165 ATTGGCCACTGTGGGGCAGGGGTGTGACTTGGCTTCCAGCTGGAAGATTATCTCAC 2224
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2225 CCAGGCCCTAGCTATATAA-CGGGCTGTGTGAGGGGCTCCAGGGCCAGTTCACGGGG 2283
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 2284 TTCATCCCAAGACAGAAAAACATAGA 2310
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 207 TTCCTTC-CACACAGAAAAACATACA 232
Db ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 7
ABZ24607
ID ABZ24607 standard; DNA; 7280 BP.
XX
XX AC ABZ24607;
XX
XX 07-APR-2003 (first entry)
XX
XX Mouse gastric antrum mucosal protein-18 gene.
XX
XX Mouse; gastrophine; AMP-18; gastric antrum mucosal protein; mitogen;
KW growth factor; vulnery; antiulcer; gene; ds.
XX
XX Mus sp.
XX
XX Key Location/Qualifiers
FH promoter 1..1874
FT /*tag= a
FT 1945..6542
FT /*tag= b
FT /product= "Mouse pre-AMP-18"
FT /note= "contains introns"
FT 1906..1956
FT exon
FT /*tag= c
```

```
FT intron /number= 1
FT 1957..3531
FT /*tag= d
FT /number= 1
FT 3532..3582
FT /*tag= e
FT /number= 2
FT 3583..3672
FT /*tag= f
FT /number= 2
FT 3673..3813
FT /*tag= g
FT /number= 3
FT 3814..4594
FT /*tag= h
FT /number= 3
FT 4595..4705
FT /*tag= i
FT /number= 4
FT 4706..5607
FT /*tag= j
FT /number= 4
FT 5608..5749
FT /*tag= k
FT /number= 5
FT 5750..6444
FT /*tag= l
FT /number= 5
FT 6445..6542
FT /*tag= m
FT /number= 6
FT 6636..6641
FT /*tag= n
FT polyA_signal
XX
XX WO200292758-A2.
XX
XX 21-NOV-2002.
XX
XX 29-MAR-2002; 2002WO-US10148.
XX
XX 29-MAR-2001; 2001US-0821726.
XX
XX (UYCH-) UNIV CHICAGO.
XX
XX Toback FG, Martin TE, Powell CT, Agarwal K;
XX WPI; 2003-120666/11.
XX P-PSDB; ABP58258.
XX
XX Gastric Antrum Mucosal Protein 18, useful for preparing a composition
XX for healing of the injured gastrointestinal tract.
XX
XX Disclosure; Fig 4; 67pp; English.
XX
XX The present sequence is that of a gene encoding a novel member of
XX the murine gastrophine group, designated antrum mucosal protein 18
XX (AMP-18). AMP-18 was initially detected in mammalian gastric
XX antrum mucosa by a differential screen of cDNA libraries obtained
XX from different regions of the pig stomach. A cDNA was also
XX isolated from a mouse library (see ABZ24608). Genomic AMP-18 DNA
XX sequences were subsequently cloned as a prelude to the analysis of
XX gene regulatory elements. AMP-18 protein is expressed at high
XX levels only in the gastric antrum, and is synthesised in the
XX luminal surface mucosal cells. Partially purified AMP-18s from
XX mouse and pig antrum tissue are mitogenic to confluent stomach
XX and kidney epithelial cells in culture. This effect is inhibited
XX by specific antisera, suggesting that AMP-18, or its products, is a
XX growth factor. AMP-18 may play a critical role in the repair of
XX the stomach epithelium following damage by agents such as alcohol,
XX non-steroidal antiinflammatory drugs, or pathogens, particularly
XX Helicobacter pylori, a causative agent of gastric ulcers and
XX possibly cancers. AMP-18 protein, or an active peptide of AMP-18,
XX is used in a claimed method of stimulating the growth of epithelial
```

```
CC cells in the gastrointestinal tract.
XX
SQ Sequence 7280 BP; 2131 A; 1597 C; 1606 G; 1944 T; 2 other;

Query Match 4.0%; Score 95; DB 25; Length 7280;
Best Local Similarity 80.5%; Pred. No. 1.6e-16;
Matches 136; Conservative 0; Mismatches 30; Indels 3; Gaps 2;

QY 198 TCTAGGGCAGCAGTCTCAACCTGGGGCCCTCGACCCCTTTGGGGAATCAACGACCCCT 257
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 391 TCTAGAGCAGTGGTTCTATCCTGTGGCCATGAGCCCTTTGGGGGGTTGAACGACCCCT 450
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 258 TT-ACAGGGGTACATATCATCTATATGTCTAGGTATTTACATTTACGATTTCGTAAC 316
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 451 TTCACAGGGGTACATATCAGATATCCTGCATCTTAGCTATTACATTTATGATTTCATAAC 510
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 317 AGTAGCAAAATTACAGGTATGAATAAGCAATGAATAAATTTATGATTG 365
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 511 AGTAGCAAAATT--AGTTAGGAAGTAGGAACAAATAACGTTATGTTG 557
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 8
AAB50815
ID AAB50815 standard; DNA; 7280 BP.
XX
AC AAB50815;
XX
DT 02-APR-2003 (first entry)
XX
DE Mouse pre-AMP-18 gene.
XX
KW Cellular growth stimulating protein; gastric antrum mucosal protein;
KW gastrokine; AMP-18 protein; gastro-intestinal disorder; cell therapy;
KW ulcer; mouse; ds.
XX
OS Mus sp.

XX Key Location/Qualifiers
FH exon 1906..1956
FT /*tag= a
FT /number= 1
FT 1957..3531
FT /*tag= b
FT /number= 1
FT 3532..3582
FT /*tag= c
FT /number= 2
FT 3583..3672
FT /*tag= d
FT /number= 2
FT 3673..3813
FT /*tag= e
FT /number= 3
FT 3814..4594
FT /*tag= f
FT /number= 3
FT 4595..4705
FT /*tag= g
FT /number= 4
FT 4706..5607
FT /*tag= h
FT /number= 4
FT 5608..5749
FT /*tag= i
FT /number= 5
FT 5750..6444
FT /*tag= j
FT /number= 5
FT 6445..6542
FT /*tag= k
FT /number= 6
FT 6636
FT /*tag= l

polyA_site
```

```
XX WO200278640-A2.
XX
PD 10-OCT-2002.
XX
PF 29-MAR-2002; 2002WO-US09885.
XX
PR 29-MAR-2001; 2001US-0821726.
XX
PA (UYCH-) UNIV CHICAGO.
XX
PI Toback GF, Martin TE, Walsh-Reitz M;
XX WPI; 2003-1032339/09.
XX
PT Protein inhibitor, useful for treating gastrointestinal disorders or
PT diseases comprises isolated homologous cellular stimulating proteins -
XX Disclosure; Fig 4; 84pp; English.
XX
CC The invention relates to a protein inhibitor which comprises homologous
CC cellular growth stimulating proteins designated gastrokines. The
CC invention also provides gastric antrum mucosal proteins designated
CC AMP-18 which belongs to the novel group of gastrokines and nucleic acid
CC molecules encoding such proteins. Pharmaceutical composition comprising
CC growth stimulating peptide derived from a gastrokine protein is useful
CC for treating gastro-intestinal disorder or diseases associated with
CC overgrowth of gastric epithelia e.g. ulcer. The invention is useful in
CC cell therapy. The present sequence is mouse pre-AMP-18 gene.
XX
SQ Sequence 7280 BP; 2131 A; 1597 C; 1606 G; 1944 T; 2 other;

Query Match 4.0%; Score 95; DB 25; Length 7280;
Best Local Similarity 80.5%; Pred. No. 1.6e-16;
Matches 136; Conservative 0; Mismatches 30; Indels 3; Gaps 2;

QY 198 TCTAGGGCAGCAGTCTCAACCTGGGGCCCTCGACCCCTTTGGGGAATCAACGACCCCT 257
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 391 TCTAGAGCAGTGGTTCTATCCTGTGGCCATGAGCCCTTTGGGGGGTTGAACGACCCCT 450
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 258 TT-ACAGGGGTACATATCATCTATATGTCTAGGTATTTACATTTACGATTTCGTAAC 316
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 451 TTCACAGGGGTACATATCAGATATCCTGCATCTTAGCTATTACATTTATGATTTCATAAC 510
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 317 AGTAGCAAAATTACAGGTATGAATAAGCAATGAATAAATTTATGATTG 365
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 511 AGTAGCAAAATT--AGTTAGGAAGTAGGAACAAATAACGTTATGTTG 557
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 9
ABZ82337/c
ID ABZ82337 standard; DNA; 18105 BP.
XX
AC ABZ82337;
XX
DT 15-MAY-2003 (first entry)
XX
DE Mouse histamine receptor H3 gene.
XX
KW Mouse; drug composition; food intake disorder; body weight disorder;
KW histamine receptor H3; anorectic; gene; ds.
XX
OS Mus musculus.
XX
FN WO2003004637-A1.
XX
PD 16-JAN-2003.
XX
PF 28-JUN-2002; 2002WO-JP06580.
XX
PR 02-JUL-2001; 2001JP-0201413.
XX
PA (BANY ) BANYU PHARM CO LTD.
```


CC mammals containing a DNA expression cassette comprising the neuron-specific promoter are useful as laboratory models for studying the function of the Presenilin gene, and for studying the etiology of CC Alzheimer's disease.

XX SQ Sequence 48974 BP; 12579 A; 10749 C; 11189 G; 14438 T; 19 other;

Query Match 3.8%; Score 88.6; DB 20; Length 48974;
Best Local Similarity 59.2%; Pred. No. 3.5e-14;
Matches 135; Conservative 0; Mismatches 59; Indels 1; Gaps 1;

QY 191 GCTGATTCTAGGCGCAGCAGTCTCAACCTGGGGGCTCGACCCCTTTGGGGGAATCAAA 250
DB 15385 GCTACACTCTAGATCGGTGTTCTCAAGCTGGGGTCCGACCCCTTTGGGGTTCGACA 15326

QY 251 CGACCCCTTTACAGGGGTGCATATCATCTATCTATATGTCAGGTATTTACATTACGATT 310
DB 15325 A-CCTTTTCACAGGGTGCACATCAGAGATCTCGCATATCCGATATTTATGTTATGATT 15267

QY 311 CGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAAATAATTTATGATTGAAGGT 370
DB 15266 CATAACAGCAACAGAAATTACAGTTAGGAAGTAGAAATGAAATAATGTTATGATTGGGC 15207

QY 371 CACCACAACATGAGG 385
DB 15206 CACCATGACATGAAG 15192

RESULT 12
ABT09109
ID ABT09109 standard; DNA; 707 BP.
XX
AC ABT09109;
XX
DT 05-DEC-2002 (first entry)
DE
DE Phase-1 Rat CT gene SEQ ID No 197.
XX
XX Rat; toxicity study; rat toxic response gene; toxicological response;
XX drug development; phase-1 rat CT gene; ds.
XX
XX Rattus sp.
XX
XX WO200266682-A2.
XX
XX 29-AUG-2002.
XX
XX 29-JAN-2002; 2002WO-US02935.
XX
XX 29-JAN-2001; 2001US-264933P.
XX
XX 26-JUL-2001; 2001US-308161P.
XX
XX (PHAS-) PHASE-1 MOLECULAR TOXICOLOGY INC.
XX
XX Farris G, Hicken SH, Farr SB;
XX
XX WPI; 2002-674961/72.
XX
XX Evaluating the toxicity of an agent, useful in drug development or in
XX determining toxicological responses to a new drug, by determining the
XX expression of rat toxicologically relevant genes in the test animal in
XX response to the test agent -
XX
XX Disclosure; Page 155; 388pp; English.
XX
XX The invention relates to a method used for evaluating the toxicity of an
XX agent comprising determining the expression of a rat toxic response
XX gene(s) in the test animal in response to the agent. The method is useful
XX in drug development, particularly for conducting toxicity studies and
XX analysis before a new drug or compound is approved for human consumption
XX or use. The method is also useful in determining toxicological responses
XX to a new drug. This polynucleotide sequence represents a phase-1 rat CT
XX gene of the invention.

XX SQ Sequence 707 BP; 205 A; 174 C; 156 G; 169 T; 3 other;

Query Match 3.5%; Score 83.6; DB 24; Length 707;
Best Local Similarity 67.3%; Pred. No. 9.9e-14;
Matches 148; Conservative 0; Mismatches 69; Indels 3; Gaps 2;

QY 171 GAGCTTCTCTTTCAGGATCAGCCTGATTCTAGGGCAGCAGTCTCAACCTGGGGGCTCG 230
DB 232 GAGCATCTCCTCAGCCCGAGCCTGTTTATGGAAGTATCTCACTCATGGGTCTATG 291

QY 231 ACCCTCTTTGGGGG--AATCAAAACGACCCCTTTTACAGGGGTCAATATCATCTATCTATAT 288
DB 292 ACCCTCTTTGGGGTTAAATGACCCCTTTCACATATCAAAATCAAAATACCTTCGAG 351

QY 289 GTCAGGATTTACATTCAGATTTCGTAACAGTAGCAAAATACAGGTATGAAATAGCAATG 348
DB 352 AGCATATATTCATTTGCAATCCGCAACAGCAGCAAAATACAGTTACGAAGTAGCAAG 411

QY 349 -AAATAATTTTATGATTGAAGGTCAACCAACATGAGGCC 387
DB 412 AAATAATCTTACGTTTGAAGGTCAATCAACACACGAGGAC 451

RESULT 13
ABX16390/C
ID ABX16390 standard; DNA; 659158 BP.
XX
AC ABX16390;
XX
DT 09-APR-2003 (first entry)
XX
DE Mouse high growth region.
XX
XX High growth region; high growth phenotype; Socs2; body size;
XX suppressor of cytokine signaling 2; ds; mouse.
XX
XX Mus sp.
XX
XX US2002155564-A1.
XX
XX 24-OCT-2002.
XX
XX 26-JAN-2001; 2001US-0771208.
XX
XX 29-DEC-1997; 97US-0999477.
XX
XX (REGC) UNIV CALIFORNIA.
XX
XX Medrano JF, Bradford E, Horvat S;
XX
XX WPI; 2003-182637/18.
XX
XX Novel gene that when downregulated or knocked-out, results in high
XX growth phenotype, useful for regulating body size in mammals e.g.
XX rodent, bovine and canine -
XX
XX Disclosure; SEQ ID NO 18; 49pp; English.
XX
XX The invention describes an isolated nucleic acid molecule encoding a
XX gene product that, when knocked out, results in a high growth (hg)
XX phenotype. For example a nucleic acid disrupting the Socs2 gene is
XX useful for producing an animal characterised by a hg phenotype, by
XX inhibiting expression of Socs2 (suppressor of cytokine signaling 2)
XX gene. The nucleic acids of the invention are useful for regulating
XX body size in mammals. gene. The nucleic acids of the invention
XX are useful for regulating body size in mammals. This sequence
XX represents the mouse high growth region.
XX Note: this sequence did not form part of the printed specification
XX but was obtained in electronic format directly from the US patent
XX office at
XX seqdata.uspto.gov/sequence.html?DocID=20020155564.
XX

SQ Sequence 659158 BP; 177521 A; 147222 C; 149414 G; 184754 T; 247 other;
Query Match 3.5%; Score 82; DB 25; Length 659158;
Best Local Similarity 71.8%; Pred. No. 1.3e-11;
Matches 135; Conservative 0; Mismatches 50; Indels 3; Gaps 2;
QY 198 TCTAGGGCAGCAGTTCTCAACCTGGGGGCTCGACCCCTTTGGGGAATCAACGACCT 257
DB 439418 TCTAGACTAGTGGTCTCAACCTATGCGTTTCAACCTCTTGGGAGTTTCATATCAGATA 439359
QY 258 TTACAGGGGTACATATCATCTATCTATATGTCAGGTATTTACATTCAGATTGCTAACA 317
DB 439358 TT--CTGCGTAGCATATCAGACATCTCTGCATATCAGATATTTCCATATGACACATATG 439301
QY 318 GTAGCAAAATTTACAGGTATGAAATAGCAATAGCAATAAATTTATGATGTAAGGTCAACCA 377
DB 439300 GTAGCAAAATTTACAGGTATGAGTAAATGAAATGATTTATGTTG--GGGTACCGCA 439242
QY 378 ACATGAGG 385
DB 439241 ACATGTGG 439234

RESULT 14
ABA97708
ID ABA97708 standard; DNA; 10917 BP.
XX
AC ABA97708;
DT 18-JUN-2002 (first entry)
XX
DE Wild-type mouse Wolfram Syndrome 1 gene.
XX
KW Wolfram Syndrome 1; WFS1; transgenic; wolframin; brain; depression;
KW stress-induced neurochemical change; behavioural change; mouse;
KW drug screening; antidepressant; gene; ds.
XX
OS Mus musculus.
XX
PN WO200191548-A2.
XX
PD 06-DEC-2001.
XX
PF 31-MAY-2001; 2001WO-US17652.
XX
PR 01-JUN-2000; 2000US-209394P.
XX
PA (PHAA) PHARMACIA & UPJOHN CO.
XX
PI Roberds SL, Huff RM;
XX
DR WPI; 2002-089969/12.
XX
PT New transgenic non-human animals (mice), useful as models for
PT depression, for monitoring the efficacy of a drug against depression,
PT and for screening antidepressants, drugs or genes for ameliorating or
PT treating depression -
XX
PS Claim 19; Page 60-63; 65pp; English.

CC The present sequence represents the genomic DNA sequence of the wild-type
CC mouse Wolfram Syndrome 1 (WFS1) gene used in the method of the invention.
CC The specification describes a non-human transgenic mammal comprising a
CC genome containing a single copy of a wild-type WFS1 gene and a modified
CC WFS1 allele containing a mutation that disrupts the function of
CC wolframin, or its transgenic progeny. Wolframin is a predicted
CC transmembrane protein expressed in many tissues including pancreas and
CC brain. The transgenic nonhuman animals are useful as models for
CC depression, particularly as models for depression in humans. The animals
CC exhibit chronic stress-induced neurochemical and behavioural changes
CC associated with depression. The transgenic animals are also useful for
CC screening or identifying antidepressant agents, drugs or genes that may
CC be employed to ameliorate or treat depression. The animals are also

CC useful for monitoring the efficacy of a drug against depression.
XX
SQ Sequence 10917 BP; 2423 A; 2842 C; 2929 G; 2706 T; 17 other;
Query Match 3.4%; Score 80.4; DB 24; Length 10917;
Best Local Similarity 70.8%; Pred. No. 3.9e-12;
Matches 136; Conservative 0; Mismatches 51; Indels 5; Gaps 2;
QY 197 TTCTAGGGCAGCAGTTCTCAACCTGGGGGCTCGACCCCTTTGGGGAATCAACGACCC 256
DB 10340 TCGGAGGACAGAGTTCTCAACCTGTGGGACACACCCCTTTGGG--TTTGAATGACIT 10397
QY 257 TTACAGGGGTCAATATCA---TCTATCTATATGTCAGGTATTTACATTCAGATTCGT 313
DB 10398 TTACAGGGGTTCGTTGCCAAAGACCATCAGAAAAACACAGGTATTTACATTCCAATTCAT 10457
QY 314 ACAGTACCAAAATTTACAGGTATGAAATAGCAATGAAATAATTTATGATTCGAGGTAC 373
DB 10458 GACAGTAGTAAATTTACAGTTATGCAAGTATGCAAGTAAATTTATGTTGGGTGAC 10517
QY 374 CACAACATGAGG 385
DB 10518 CACATCATGAG 10529
RESULT 15
AAV06155
ID AAV06155 standard; DNA; 892 BP.
XX
AC AAV06155;
XX
DT 06-MAY-1998 (first entry)
XX
DE Viral infection gene SEQ ID NO:75.
XX
KW Viral infection; tumour suppressor; cellular gene; rat; cancer;
KW serum protein; inhibitor; malignant phenotype; HIV; influenza;
KW hepatitis; retrovirus; immunodeficiency; ds.
XX
OS Rattus sp.
XX
PN WO9739119-A1.
XX
PD 23-OCT-1997.
XX
PF 11-APR-1997; 97WO-US06067.
XX
PR 15-APR-1996; 96US-0015334.
XX
PA (UYVA-) UNIV VANDERBILT.
XX
PI Dubois RN, Organ EL, Rubin DH;
XX
DR WPI; 1997-526456/48.
XX
PT Genes involved in viral infection and tumour suppression - used to
PT develop products for reducing or preventing viral infection or for
PT suppressing tumours
XX
PS Claim 1; Page 82-83; 101pp; English.
XX
CC The present sequence represents a viral infection gene. The present
CC invention describes nucleic acid sequences isolated from rat. The
CC sequences of the invention comprise 70 viral infection (VI) genes and
CC 8 tumour suppressor (TS) genes. Propagating cell cultures in the
CC absence of the serum protein (SP) allows selective elimination of cells
CC persistently infected with a virus from the cell culture. Inhibitors of
CC the SP can be used for reducing or inhibiting a viral infection.
CC Inhibitors of the TS gene products can be used to suppress a malignant
CC phenotype (MP). The methods and inhibitors can be used with viruses
CC such as HIV, influenza, hepatitis virus or animal retroviruses such as
CC simian immunodeficiency virus, avian immunodeficiency virus, equine
CC immunodeficiency virus, feline immunodeficiency virus, equine infectious

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 23:31:50 ; Search time 137.266 Seconds
(without alignments)
7582.209 Million cell updates/sec

Title: US-10-005-337A-1
Perfect score: 2358
Sequence: 1 ggcattcttcattttaaca.....caggtcggaggccaccatgg 2358

Scoring table: IDENTITY NUC
Gapop 10_0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA: *
1: /cgn2.6/prodata/2/ina/5A_COMB.seq: *
2: /cgn2.6/prodata/2/ina/5B_COMB.seq: *
3: /cgn2.6/prodata/2/ina/6A_COMB.seq: *
4: /cgn2.6/prodata/2/ina/6B_COMB.seq: *
5: /cgn2.6/prodata/2/ina/PCUS_COMB.seq: *
6: /cgn2.6/prodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	88.6	3.8	48974	3	US-08-920-422-17
C 2	78.8	3.3	892	4	US-09-171-209-75
C 3	74.8	3.2	37950	3	US-09-338-907-183
C 4	74.8	3.2	37950	4	US-09-218-207-183
C 5	74.6	3.2	10614	1	US-08-135-511-35
C 6	74.6	3.2	10614	1	US-08-187-453-35
C 7	73.4	3.1	4164	4	US-08-882-164-38
C 8	68.6	2.9	4072	3	US-09-272-496-7
C 9	68	2.9	16442	3	US-08-781-891-208
C 10	68	2.9	16442	4	US-09-618-166-208
C 11	65.6	2.8	29604	3	US-08-781-891-207
C 12	65.6	2.8	29604	4	US-09-618-166-207
C 13	65	2.8	6645	2	US-08-380-403A-4
C 14	65	2.8	6645	2	US-08-895-628-4
C 15	65	2.8	6645	4	US-08-895-810D-4
C 16	63	2.7	3450	2	US-08-378-617A-9
C 17	62	2.6	7208	3	US-09-166-186-107
C 18	62	2.6	7208	3	US-09-313-932-107
C 19	54	2.3	7218	1	US-08-232-463-14
C 20	50.6	2.1	5764	4	US-09-312-762A-8
C 21	47	2.0	26700	1	US-08-472-217-1
C 22	47	2.0	26700	2	US-08-488-199-5
C 23	47	2.0	26700	3	US-08-760-534A-1
C 24	47	2.0	26700	4	US-09-336-757-1
C 25	38.8	1.6	818	4	US-09-328-475C-157
C 26	38.6	1.6	2830	1	US-08-604-333-3
C 27	38.6	1.6	2830	3	US-09-110-618-3
C 17	62	2.6	7208	3	US-09-166-186-107
C 18	62	2.6	7208	3	US-09-313-932-107
C 19	54	2.3	7218	1	US-08-232-463-14
C 20	50.6	2.1	5764	4	US-09-312-762A-8
C 21	47	2.0	26700	1	US-08-472-217-1
C 22	47	2.0	26700	2	US-08-488-199-5
C 23	47	2.0	26700	3	US-08-760-534A-1
C 24	47	2.0	26700	4	US-09-336-757-1
C 25	38.8	1.6	818	4	US-09-328-475C-157
C 26	38.6	1.6	2830	1	US-08-604-333-3
C 27	38.6	1.6	2830	3	US-09-110-618-3

28	38.6	1.6	2830	4	US-09-578-178-3	Sequence 3, Appli
29	38.6	1.6	2830	4	US-09-577-806-3	Sequence 3, Appli
30	37.2	1.6	7218	1	US-08-232-463-14	Sequence 14, Appli
31	36.8	1.6	277	3	US-09-007-005-3	Sequence 3, Appli
32	36.8	1.6	277	3	US-09-244-796-3	Sequence 3, Appli
33	36.6	1.6	1664976	4	US-08-916-421B-1	Sequence 1, Appli
34	36.2	1.5	1209	4	US-08-462-509B-1	Sequence 1, Appli
35	36.2	1.5	1209	5	PCT-US95-05616-1	Sequence 1, Appli
36	36.2	1.5	1564	2	US-08-846-705-4	Sequence 4, Appli
37	36.2	1.5	1564	3	US-08-846-704-1	Sequence 1, Appli
38	36.2	1.5	1564	3	US-08-846-704-3	Sequence 3, Appli
39	35	1.5	9573	4	US-09-220-132-168	Sequence 168, App
C 40	35	1.5	80346	3	US-09-078-294-4	Sequence 4, Appli
C 41	35	1.5	80595	3	US-09-078-294-3	Sequence 3, Appli
C 42	34.8	1.5	2088	2	US-08-602-264A-1	Sequence 1, Appli
C 43	34.8	1.5	2088	3	US-08-461-018A-1	Sequence 1, Appli
C 44	34.8	1.5	2088	3	US-09-216-958-1	Sequence 1, Appli
C 45	34.4	1.5	1843	4	US-09-622-540A-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1
US-08-920-422-17/c
; Sequence 17, Application US/08920422A
; Patent No. 6255473
; GENERAL INFORMATION:
; APPLICANT: Vitek, Michael P.
; APPLICANT: Mitsuda, No. 6255473iaki
; APPLICANT: Roses, Allen D.
; TITLE OF INVENTION: Presenilin-1 Gene Promoter
; FILE REFERENCE: VITEKPRESENILIN
; CURRENT APPLICATION NUMBER: US/08/920,422A
; CURRENT FILING DATE: 1997-08-29
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 17
; LENGTH: 48974
; TYPE: DNA
; ORGANISM: Mus musculus
US-08-920-422-17

Query Match 3.8%; Score 88.6; DB 3; Length 48974;
Best Local Similarity 69.2%; Pred. No. 3 9e-16;
Matches 135; Conservative 0; Mismatches 59; Indels 1; Gaps 1;
QY 191 GCCTGATTTAGGGCAGCAGTTCTCAACCTGGGGCCTCGACCCCTTTGGGGGAATCAAA 250
DB 15385 GCTACACTTAGATCGGTGGTTCTCAAGCTGGGGTCCGCGACCCCTTTGGGGTGGACA 15326
QY 251 CGACCCCTTACAGGGGTACATATCATCTATATGTCTAGGTATTTACATTACGATT 310
DB 15325 A-CCTTTTTCACAAAGGTTCACATCAGAGATCCTGCATATCCGATATTTATGTTATGTT 15267
QY 311 CTTAACAGTAGCAAAATACAGGTATGAATAGCAATGAATAATTTATGTTGAAGT 370
DB 15266 CATACAGCAACAGAAATACAGTTAGGAAGTAGAATAATATGTTATGTTGGGC 15207
QY 371 CACCAACAACATGAGG 385
DB 15206 CACCATGACATGAAG 15192
RESULT 2
US-09-171-209-75
; Sequence 75, Application US/09171209
; Patent No. 6448000
; GENERAL INFORMATION:
; APPLICANT: VANDERBILT UNIVERSITY
; APPLICANT: 305 Kirkland Hall
; Nashville, TN 37240
; TITLE OF INVENTION: MAMMALIAN GENES INVOLVED IN VIRAL


```

; APPLICANT: Yu, Chang-En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/09/618,166
; FILING DATE: 17-Jul-2000
; CLASSIFICATION: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: McMasters, David D.
; REGISTRATION NUMBER: 33,963
; REFERENCE/DOCKET NUMBER: 240052.419C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 208:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16442 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; SEQUENCE DESCRIPTION: SEQ ID NO: 208:
US-09-618-166-208

Query Match 2.9%; Score 68; DB 3; Length 16442;
Best Local Similarity 66.2%; Pred. No. 5e-10;
Matches 98; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 238 TGGGGGAATCAACGACCCCTTTACAGGGTCACATATCATCTATCTATATGTGAGGTAT 297
Db 6043 TGATTGATCAAAATGGCTCTTTACAGGGTTTACAAATCAACATCCTGTATATCAGATAG 5984

QY 298 TTACATTACGATTCGTAACAGTAGCAAAATTTACAGGTATCAATGCAATGAATAATTT 357
Db 5983 TTACATCAAAATTCATAACGGTGGCAAAATTTACAGTCAAGAAGTAGCAAGAAATAATG 5924

QY 358 TATGATTGAAGGTCCACCAACATGAGG 385
Db 5923 TCATGTTGGTGCCATCAGCAGCAGG 5896

RESULT 10
US-09-618-166-208/c
; Sequence 208, Application US/09618166
; Patent No. 6583112
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; Yu, Chang-En
; Oshima, Junko
; Mulligan, John T.
; Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed Intellectual Property Law Group
; STREET: 701 Fifth Avenue, Suite 6300
; CITY: Seattle
; STATE: Washington

; APPLICANT: Yu, Chang-En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: No. 6090820tenburg Ph.D., Carol
; REGISTRATION NUMBER: 39,317
; REFERENCE/DOCKET NUMBER: 240052.419
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 208:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16442 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-781-891-208

Query Match 2.9%; Score 68; DB 3; Length 16442;
Best Local Similarity 66.2%; Pred. No. 5e-10;
Matches 98; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 238 TGGGGGAATCAACGACCCCTTTACAGGGTCACATATCATCTATCTATATGTGAGGTAT 297
Db 6043 TGATTGATCAAAATGGCTCTTTACAGGGTTTACAAATCAACATCCTGTATATCAGATAG 5984

QY 298 TTACATTACGATTCGTAACAGTAGCAAAATTTACAGGTATCAATGCAATGAATAATTT 357
Db 5983 TTACATCAAAATTCATAACGGTGGCAAAATTTACAGTCAAGAAGTAGCAAGAAATAATG 5924

QY 358 TATGATTGAAGGTCCACCAACATGAGG 385
Db 5923 TCATGTTGGTGCCATCAGCAGCAGG 5896

RESULT 11
US-08-781-891-207
; Sequence 207, Application US/08781891
; Patent No. 6090620
; GENERAL INFORMATION:
; APPLICANT: Fu, Ying-Hui
; APPLICANT: Yu, Chang-En
; APPLICANT: Oshima, Junko
; APPLICANT: Mulligan, John T.
; APPLICANT: Schellenberg, Gerald D.
; TITLE OF INVENTION: GENE AND GENE PRODUCTS RELATED TO
; TITLE OF INVENTION: WERNER'S SYNDROME
; NUMBER OF SEQUENCES: 209
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED and BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/781,891
; FILING DATE: 27-DEC-1996
; CLASSIFICATION: 800
```



```
Best Local Similarity 72.3%; Pred. No. 2.5e-09;
Matches 112; Conservative 0; Mismatches 40; Indels 3; Gaps 2;

QY 233 CCCTTTGGGGGAATCAACGACCTTTACAGGGGTCCACAT--ATCATCTATCTCTATATGT 290
Db 554 CCCTTCTGAGGGTCTAAACGGCCCATTCACAGGGGTCCGATGCAGCAGACACCTGTGTGT 495
QY 291 CAGGTATTACATTACGATTTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAA 350
Db 494 CAGGTGTTTATGTGTACAAATTCCTTAAGGGTAGCAGAAATTACAGTTATGAAGTGCATGAA 435
QY 351 ATAATTTTATGATTGAAGGTCAACCAACATGAGG 385
Db 434 AATAATTTTGA-TCAGGGTCAACCAACATGTGG 401

RESULT 14
US-08-895-628-4/c
; Sequence 4, Application US/08895628
; Patent No. 5998585
; GENERAL INFORMATION:
; APPLICANT: MINATO, Nagahiro
; APPLICANT: HATTORI, Masakazu
; APPLICANT: HIROSHI, Kubota
; APPLICANT: MASATSUGU, Maeda
; TITLE OF INVENTION: SPA-1 PROTEIN AND GENE CODING THEREFOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/895,628
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/380,403
; FILING DATE: 30-JAN-1995
; APPLICATION NUMBER: US 08/325,909
; FILING DATE: 19-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 6-279712
; FILING DATE: 20-OCT-1994
; APPLICATION NUMBER: JP 6-139513
; FILING DATE: 30-MAY-1994
; NAME: WEGNER, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 53466/128/AAOK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6645 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Join(904..1015, 1356..1459, 1726..1883, 2009
; LOCATION: ..2618, 2890..3164, 4291..4509, 4598..4709, 4795

;
; LOCATION: ..4903, 5017..5117, 5200..5255, 5447..5525, 5598
; LOCATION: ..5741)
; US-08-895-628-4
Query Match 2.8%; Score 65; DB 2; Length 6645;
Best Local Similarity 72.3%; Pred. No. 2.5e-09;
Matches 112; Conservative 0; Mismatches 40; Indels 3; Gaps 2;

QY 233 CCCTTTGGGGGAATCAACGACCTTTACAGGGGTCCACAT--ATCATCTATCTCTATATGT 290
Db 554 CCCTTCTGAGGGTCTAAACGGCCCATTCACAGGGGTCCGATGCAGCAGACACCTGTGTGT 495
QY 291 CAGGTATTACATTACGATTTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAA 350
Db 494 CAGGTGTTTATGTGTACAAATTCCTTAAGGGTAGCAGAAATTACAGTTATGAAGTGCATGAA 435
QY 351 ATAATTTTATGATTGAAGGTCAACCAACATGAGG 385
Db 434 AATAATTTTGA-TCAGGGTCAACCAACATGTGG 401

RESULT 15
US-08-895-810D-4/c
; Sequence 4, Application US/08895810D
; Patent No. 6406886
; GENERAL INFORMATION:
; APPLICANT: MINATO, Nagahiro
; APPLICANT: HATTORI, Masakazu
; APPLICANT: HIROSHI, Kubota
; APPLICANT: MASATSUGU, Maeda
; TITLE OF INVENTION: SPA-1 PROTEIN AND GENE CODING THEREFOR
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/895,810D
; FILING DATE: 17-Jul-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/380,403
; FILING DATE: 30-JAN-1995
; APPLICATION NUMBER: US 08/325,909
; FILING DATE: 19-OCT-1994
; APPLICATION NUMBER: JP 6-279712
; FILING DATE: 20-OCT-1994
; APPLICATION NUMBER: JP 6-139513
; FILING DATE: 30-MAY-1994
; NAME: WEGNER, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 53466/128/AAOK
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6645 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
```

```

; NAME/KEY: CDS
; LOCATION: join(904..1015, 1356..1459, 1726..1883, 2009
; .2618, 2890..3164, 4291..4509, 4598..4709, 4795..4903,
; 5017..5117, 5200..5255, 5447..5525, 5598..5741)
; SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-08-895-810D-4

```

```

Query Match      2.8%; Score 65; DB 4; Length 6645;
Best Local Similarity 72.3%; Pred. No. 2.5e-09;
Matches 112; Conservative 0; Mismatches 40; Indels 3; Gaps 2;

QY 233 CCCTTTGGGGGAATCAAAACGACCTTTACAGGGGTACACAT--ATCATCTATCCTATATGT 290
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 554 CCCTTCTGAGGGTCTAACGGCCCATTCACAGGGGTGCGATGCACAGACCCCTGTGTGT 495

QY 291 CAGGTATTTACATTACGATTCGTAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAA 350
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 494 CAGGTGTTTATGTTACAAATTCCTAAGGGTAGCAGAAATTACAGTTATGAAAGTGGCAATGAA 435

QY 351 ATAATTTTATGATTGAAGGTCACCACACACATGAGG 385
    |||||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||  ||
Db 434 ATAATTTTGA-TCAGGGTCACCACCATGTGG 401

```

Search completed: November 18, 2003, 07:06:16
Job time : 147.266 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 23:33:10 ; Search time 1128.46 Seconds
(without alignments)
6829.176 Million cell updates/sec

Title: US-10-005-337A-1

Perfect score: 2358

Sequence: 1 ggatccttcatgtttaaca.....caggtcggagccaccatgg 2358

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 2169961 seqs, 1634102185 residues

Total number of hits satisfying chosen parameters: 433922

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA:

- 1: /cgn2_6/prodata/1/pubpna/US07_PUBCOMB.seq:
- 2: /cgn2_6/prodata/1/pubpna/PCT_NEW_PUB.seq:
- 3: /cgn2_6/prodata/1/pubpna/US06_NEW_PUB.seq:
- 4: /cgn2_6/prodata/1/pubpna/US06_PUBCOMB.seq:
- 5: /cgn2_6/prodata/1/pubpna/US07_NEW_PUB.seq:
- 6: /cgn2_6/prodata/1/pubpna/PCTUS_PUBCOMB.seq:
- 7: /cgn2_6/prodata/1/pubpna/US08_NEW_PUB.seq:
- 8: /cgn2_6/prodata/1/pubpna/US08_PUBCOMB.seq:
- 9: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:
- 10: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:
- 11: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:
- 12: /cgn2_6/prodata/1/pubpna/US09_PUBCOMB.seq:
- 13: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:
- 14: /cgn2_6/prodata/1/pubpna/US10_PUBCOMB.seq:
- 15: /cgn2_6/prodata/1/pubpna/US10_NEW_PUB.seq:
- 16: /cgn2_6/prodata/1/pubpna/US60_NEW_PUB.seq:
- 17: /cgn2_6/prodata/1/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2358	100.0	2358	14	US-10-005-337A-1
2	367.8	15.6	2074	14	US-10-005-337A-2
3	100.4	4.3	90650	14	US-10-175-523-80
4	88.8	3.8	659158	10	US-09-771-208-20
5	84.2	3.6	106664	14	US-10-175-523-97
6	82	3.5	659158	10	US-09-771-208-20
7	80.6	3.4	889	12	US-10-002-631C-156
8	80.4	3.4	10917	12	US-10-195-963-1
9	78.2	3.3	185548	14	US-10-175-523-62
10	78	3.3	892	14	US-10-228-794-75
11	76.8	3.3	249487	13	US-09-728-445-91
12	75.8	3.2	335	10	US-09-917-800A-477
13	75.6	3.2	5390	10	US-09-917-800A-477
14	74.8	3.2	37950	10	US-09-901-484A-183
15	74.8	3.2	37950	10	US-09-853-526-183
16	71.6	3.0	171936	12	US-10-025-966A-24

c	17	71.6	3.0	171936	12	US-10-265-071-24	Sequence 24, Appl
	18	69.4	2.9	1889	10	US-09-974-298-182	Sequence 182, Appl
	19	69.4	2.9	1889	12	US-10-240-965-158	Sequence 158, Appl
	20	69.4	2.9	1889	12	US-10-253-157-421	Sequence 421, Appl
c	21	68.8	2.9	106664	14	US-10-175-523-97	Sequence 97, Appl
c	22	68	2.9	185548	14	US-10-175-523-62	Sequence 62, Appl
c	23	67.8	2.9	1325	14	US-10-084-817-217	Sequence 217, Appl
c	24	64	2.7	85548	14	US-10-175-523-75	Sequence 75, Appl
	25	63.8	2.7	37051	12	US-10-004-113-55	Sequence 55, Appl
	26	63.2	2.7	184	9	US-09-923-876-5218	Sequence 5218, Ap
	27	63	2.7	3240	12	US-10-125-994A-25	Sequence 25, Appl
	28	63	2.7	3537	12	US-10-125-994A-26	Sequence 26, Appl
c	29	62	2.6	7208	11	US-09-824-322B-107	Sequence 107, Appl
c	30	60.6	2.6	67191	12	US-10-105-612-1	Sequence 1, Appl
c	31	60.4	2.6	696	9	US-09-728-446-115	Sequence 115, Appl
c	32	60	2.5	32069	12	US-10-004-113-7	Sequence 7, Appl
c	33	59.8	2.5	347	9	US-09-728-448-339	Sequence 339, Appl
c	34	59.6	2.5	4109	11	US-09-866-050A-567	Sequence 567, Appl
c	35	59.6	2.5	4109	14	US-10-152-661-567	Sequence 567, Appl
c	36	59.4	2.5	520	10	US-09-867-701-1380	Sequence 1380, Ap
c	37	57.6	2.4	6000	9	US-09-738-968-44	Sequence 44, Appl
c	38	57.4	2.4	893	13	US-10-044-592-3	Sequence 3, Appl
c	39	57.4	2.4	955	13	US-10-044-592-93	Sequence 93, Appl
c	40	57.4	2.4	1574	13	US-10-044-592-95	Sequence 95, Appl
c	41	56.6	2.4	1235	10	US-09-809-545A-15	Sequence 15, Appl
c	42	56.6	2.4	2120	9	US-09-801-574-29	Sequence 29, Appl
c	43	56.2	2.4	74868	14	US-10-175-523-67	Sequence 67, Appl
c	44	55.6	2.4	1470	14	US-10-208-304-6	Sequence 6, Appl
c	45	55.6	2.4	173808	13	US-10-003-806-10	Sequence 10, Appl

ALIGNMENTS

RESULT 1

US-10-005-337A-1
; Sequence 1, Application US/10005337A
; Publication No. US20030039984A1
; GENERAL INFORMATION:
; APPLICANT: BENOIT, Patrick
; APPLICANT: SCHWARTZ, Bertrand
; APPLICANT: BRANLLEUC, Didier
; APPLICANT: CHIEN, Kenneth R.
; TITLE OF INVENTION: SEQUENCES UPSTREAM OF THE CARP GENE, VECTORS CONTAINING
; FILE OF INVENTION: THEM AND USES THEREOF
; FILE REFERENCE: 03806.0530-00000
; CURRENT APPLICATION NUMBER: US/10/005,337A
; CURRENT FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: US 60/251,582
; PRIOR FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 2358
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-005-337A-1

Query Match 100.0%; Score 2358; DB 14; Length 2358;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2358; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	GGATCCTTTCATGTTTAAACAATATCAACCTTAACCCAGGGAACAGCTGCTGACAGT	60
Db	1	GGATCCTTTCATGTTTAAACAATATCAACCTTAACCCAGGGAACAGCTGCTGACAGT	60
Qy	61	GGCTTTGGCCACCCTCAATACCTTCTAGTCTAGTCCGTTTGTGAAACTCAGCCCATCCCA	120
Db	61	GGCTTTGGCCACCCTCAATACCTTCTAGTCTAGTCCGTTTGTGAAACTCAGCCCATCCCA	120
Qy	121	ACACTTCTGCAAGCCCATCCTCTCAAGGTGCTCATTTGGGAATTTCTTGGAGCTTCTCT	180
Db	121	ACACTTCTGCAAGCCCATCCTCTCAAGGTGCTCATTTGGGAATTTCTTGGAGCTTCTCT	180

Qy	181	TTCAAGGATCAGCCCTGATTCTTAGGCACAGGTTCTCAACCTGGGGGCTCGACCCCTTTGG	240
Db	181	TTCAAGGATCAGCCCTGATTCTTAGGCACAGTTCCTCAACCTGGGGGCTCGACCCCTTTGG	240
Qy	241	GGGNAATCAAAAGCACCCTTTACAGGGGTACATATCATCTATCTCTATATGTCAAGTATTTA	300
Db	241	GGGNAATCAAAAGCACCCTTTACAGGGGTACATATCATCTATCTCTATATGTCAAGTATTTA	300
Qy	301	CATTACGATTTCGTAACAGTAGCAAAATTCACAGGTATGAAATAGCAATGAAATAAATTTAT	360
Db	301	CATTACGATTTCGTAACAGTAGCAAAATTCACAGGTATGAAATAGCAATGAAATAAATTTAT	360
Qy	361	GATTGAAGGTCAACCAACATGAGGCCGCCACACTGTTCTTAGAGAAAAATCACTGGGGTG	420
Db	361	GATTGAAGGTCAACCAACATGAGGCCGCCACACTGTTCTTAGAGAAAAATCACTGGGGTG	420
Qy	421	GGGAAAGGTTTGGAAAGCCTTTCTGTGCCATCTTTCAATCTTCAAAAGTGATGTGTTTCA	480
Db	421	GGGAAAGGTTTGGAAAGCCTTTCTGTGCCATCTTTCAATCTTCAAAAGTGATGTGTTTCA	480
Qy	481	GAAAGCCCTTCACGTGTTCTGCTGGGGCTCTTAGTAGTCTGAGTAGGAAGTCTGATGTAC	540
Db	481	GAAAGCCCTTCACGTGTTCTGCTGGGGCTCTTAGTAGTCTGAGTAGGAAGTCTGATGTAC	540
Qy	541	CAGGTCCTGCTTCTTATGGGTGGACCAAGACGCATCGTGGGTGGAGCGAAGACGCAACCT	600
Db	541	CAGGTCCTGCTTCTTATGGGTGGACCAAGACGCATCGTGGGTGGAGCGAAGACGCAACCT	600
Qy	601	CACCTTCTAGCTCTGCATCCATAGCAAGTAGCCCTAATGTTTCTGTGTCTAGGTGTCACT	660
Db	601	CACCTTCTAGCTCTGCATCCATAGCAAGTAGCCCTAATGTTTCTGTGTCTAGGTGTCACT	660
Qy	661	CTGTGAATCGAGATCCTTGGCCCTGCTTGAAATAGGGAGGCACAAATACTCAGAGATTC	720
Db	661	CTGTGAATCGAGATCCTTGGCCCTGCTTGAAATAGGGAGGCACAAATACTCAGAGATTC	720
Qy	721	AAGACTGCTCAGCAGCCAGAGTCCTTCCCTCAAAGGAAAGTGCTCAACTCTCAGCCCCC	780
Db	721	AAGACTGCTCAGCAGCCAGAGTCCTTCCCTCAAAGGAAAGTGCTCAACTCTCAGCCCCC	780
Qy	781	TTAGTCTGAGTCAGGCTCGGAACAAACGGCCACAGGAATGAGAAAAAGTGCCTAGCTG	840
Db	781	TTAGTCTGAGTCAGGCTCGGAACAAACGGCCACAGGAATGAGAAAAAGTGCCTAGCTG	840
Qy	841	CTTGTCACTTCAGAGGTCAAAGAAATAGTGTTAACCATGAAAAACGAGAAGACCAACAG	900
Db	841	CTTGTCACTTCAGAGGTCAAAGAAATAGTGTTAACCATGAAAAACGAGAAGACCAACAG	900
Qy	901	TTATCCATTGATAGCGTCTCAGGACAGATAGGACAGAGAGAACTAGGAGAGGGGAACC	960
Db	901	TTATCCATTGATAGCGTCTCAGGACAGATAGGACAGAGAGAACTAGGAGAGGGGAACC	960
Qy	961	CACGAAGGCAAAAGTATTAGTGTGTGGTTTTTCAGGGCAATGTCTTGCTAGTAAGATTCT	1020
Db	961	CACGAAGGCAAAAGTATTAGTGTGTGGTTTTTCAGGGCAATGTCTTGCTAGTAAGATTCT	1020
Qy	1021	AGAAACAAATTTGCTGGTTGAACAGCTGAAGTGGGGTGGGGTTCTTACCCCATGTTCA	1080
Db	1021	AGAAACAAATTTGCTGGTTGAACAGCTGAAGTGGGGTGGGGTTCTTACCCCATGTTCA	1080
Qy	1081	TGGAAGGGTGAGTGAGGAGAGACAGATATATGATGGCCAGCATAAACAAACATACACAACA	1140
Db	1081	TGGAAGGGTGAGTGAGGAGAGACAGATATATGATGGCCAGCATAAACAAACATACACAACA	1140
Qy	1141	CCCTAAATTAACACTTCCCTCTTCTACTGACACCCCTTCACTCTCTCTTTTCATAAAAA	1200
Db	1141	CCCTAAATTAACACTTCCCTCTTCTACTGACACCCCTTCACTCTCTCTTTTCATAAAAA	1200
Qy	1201	TAAAAAAGTATTTTATGTGGCTCTTAGCATAGNATCTTCTCTCGAATCTATAAAGATC	1260
Db	1201	TAAAAAAGTATTTTATGTGGCTCTTAGCATAGNATCTTCTCTCGAATCTATAAAGATC	1260

Qy	1261	TAAATATTTATATTTTTCACATTTTAAATATCTTTAGCGATGACAAGCCAGAGAAACAAGTATT	1320
Db	1261	TAAATATTTATATTTTTCACATTTTAAATATCTTTAGCGATGACAAGCCAGAGAAACAAGTATT	1320
Qy	1321	TTTGTGCTCTCTCAACAGCAAAAGCTTTGGGGCCCTTTTGTTCCTCGTGTAGGAATAGAACA	1380
Db	1321	TTTGTGCTCTCTCAACAGCAAAAGCTTTGGGGCCCTTTTGTTCCTCGTGTAGGAATAGAACA	1380
Qy	1381	CGAGAGCCCGTGATCTTAGGCAGATGCTCTATCATTTAGCCCATGAGTCTCCAGCCCTCAG	1440
Db	1381	CGAGAGCCCGTGATCTTAGGCAGATGCTCTATCATTTAGCCCATGAGTCTCCAGCCCTCAG	1440
Qy	1441	ACGCACATTTTTCTCGGGCTCTCTTAAAGCTTTTCCACAGCATTTGGGAAACTTTTACTGAC	1500
Db	1441	ACGCACATTTTTCTCGGGCTCTCTTAAAGCTTTTCCACAGCATTTGGGAAACTTTTACTGAC	1500
Qy	1501	AGCATCCAAGTTGTGCTTCTGCTTAAGNACTGGAGACTCACATCTCTCTGTGCACTACCTCGG	1560
Db	1501	AGCATCCAAGTTGTGCTTCTGCTTAAGNACTGGAGACTCACATCTCTCTGTGCACTACCTCGG	1560
Qy	1561	CCCGTTTTGGGTAGATCCCTCTGATTAGCCCTTCAGATTTTAGAACAACGGTGAGCCTGTGGT	1620
Db	1561	CCCGTTTTGGGTAGATCCCTCTGATTAGCCCTTCAGATTTTAGAACAACGGTGAGCCTGTGGT	1620
Qy	1621	GCATAAATTATGCCAGTGACACCATAGAGTCAAAGTGCAATTAAGTCTTCAATTT	1680
Db	1621	GCATAAATTATGCCAGTGACACCATAGAGTCAAAGTGCAATTAAGTCTTCAATTT	1680
Qy	1681	CTCCTAATGTGTTGATCGATGGCCATGTCACAGGGCCATTTTAGCTGAGACATCAGTCCAG	1740
Db	1681	CTCCTAATGTGTTGATCGATGGCCATGTCACAGGGCCATTTTAGCTGAGACATCAGTCCAG	1740
Qy	1741	AGAAATCCAAACAGATAGACACAAGTGGACCCAGACCCCATCTCTCCCTCGGGCTGA	1800
Db	1741	AGAAATCCAAACAGATAGACACAAGTGGACCCAGACCCCATCTCTCCCTCGGGCTGA	1800
Qy	1801	TTATCCCCAGAAATAGSATGTCCCAAAGCAACACTTTCCAGCCAACTGGAGTCTGTATAA	1860
Db	1801	TTATCCCCAGAAATAGSATGTCCCAAAGCAACACTTTCCAGCCAACTGGAGTCTGTATAA	1860
Qy	1861	GTCCAGTTATCAGAAAGATATGCTGTAGTGTGATGACAGTGTCTTGCATTTTCTTGAT	1920
Db	1861	GTCCAGTTATCAGAAAGATATGCTGTAGTGTGATGACAGTGTCTTGCATTTTCTTGAT	1920
Qy	1921	ACGTTAGTCATATGAGAGCTGACAAAGAGGAAAAAGACGAGCGATGTGTGTGCAATATTA	1980
Db	1921	ACGTTAGTCATATGAGAGCTGACAAAGAGGAAAAAGACGAGCGATGTGTGTGCAATATTA	1980
Qy	1981	ACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGGATGACTCGCATTTGCTGAGCGGTGTG	2040
Db	1981	ACAGGCAGCTGTCCCCTGGCTTCCCGATACGTGGGATGACTCGCATTTGCTGAGCGGTGTG	2040
Qy	2041	GTCACTGCCAAAGGAATGACCTCTCACATTTCTTCTGTGATTCGCATACGCCCGGGCCAG	2100
Db	2041	GTCACTGCCAAAGGAATGACCTCTCACATTTCTTCTGTGATTCGCATACGCCCGGGCCAG	2100
Qy	2101	CTTGTGATCTCCCTCTTGGGCTTCCAGACACTAAGTCTGGAAATGAAATTCACCTGCGCT	2160
Db	2101	CTTGTGATCTCCCTCTTGGGCTTCCAGACACTAAGTCTGGAAATGAAATTCACCTGCGCT	2160
Qy	2161	CTGAATTTGGCCACTGTGTGGGGCAGGGGTGTGACTTGGCTTCCCGAGCTGGAAAGATTATC	2220
Db	2161	CTGAATTTGGCCACTGTGTGGGGCAGGGGTGTGACTTGGCTTCCCGAGCTGGAAAGATTATC	2220
Qy	2221	TCACCCAGCCCTAGCTATATATAACGGGCTGTGTGTGAGGGGCTCCAAGGCCAGTTTCCAG	2280
Db	2221	TCACCCAGCCCTAGCTATATATAACGGGCTGTGTGTGAGGGGCTCCAAGGCCAGTTTCCAG	2280
Qy	2281	GGGTTCAATCACAAGAGAGAAAACAATAGACTCGAGGCTCTAGGAGCTTGCATGCGCTGCA	2340
Db	2281	GGGTTCAATCACAAGAGAGAAAACAATAGACTCGAGGCTCTAGGAGCTTGCATGCGCTGCA	2340
Qy	2341	GGTCGAGGCCACCATGG	2358

```
|||||
Db 2341 GGTCGGAGGCCACCATTGG 2358
|||||
RESULT 2
US-10-005-337A-2
; Sequence 2, Application US/10005337A
; Publication No. US20030039984A1
; GENERAL INFORMATION:
; APPLICANT: BENOIT, Patrick
; APPLICANT: SCHWARTZ, Bertrand
; APPLICANT: BRANELLEC, Didier
; APPLICANT: CHIEN, Kenneth R.
; TITLE OF INVENTION: SEQUENCES UPSTREAM OF THE CARP GENE, VECTORS CONTAINING
; FILE REFERENCE: 03806.0530-00000
; CURRENT APPLICATION NUMBER: US/10/005,337A
; CURRENT FILING DATE: 2001-12-07
; PRIOR APPLICATION NUMBER: US 60/251,582
; PRIOR FILING DATE: 2000-12-07
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 2074
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-005-337A-2

Query Match 15.6%; Score 367.8; DB 14; Length 2074;
Best Local Similarity 76.4%; Pred. No. 3.4e-101;
Matches 597; Conservative 0; Mismatches 162; Indels 22; Gaps 11;

QY 1540 TCTCTCTGTCATCACTTCGGCCGCTTTTGGGGT--AGATCCTCTGATTAGCTTCAGAT 1597
Db 1289 TCTATTCTTGACCACTCTGATCCATTTTGAAGTAAATGCTCCAAATTATTATGCTGT 1348
QY 1598 TTAGAACCGGTGAGCCTGTGTGCTACTAATTTATGGCCAGTGACACCATAGAGTCAAAGT 1657
Db 1349 TTAGAACCGGTGAGCCTGTGTGCTA--ATGGCCAGTGACATCATAAAGAAAGT 1405
QY 1658 GCATTACTGAATGCTTCAATTTCTCTAATCTGCTGATGATGGATGTCAGGGCCAT 1717
Db 1406 GCATTACTGAATGCTTCAATTTCTCTAATCTGCTGATGATGGATGTCAGGGCCAT 1465
QY 1718 TTTAGCTGCAGACATCACTCCAGAGAAATCCAAACAGATAGAGCAAGTGGCCACACAGAC 1777
Db 1466 TTTAGC-CAGACATCACTCCAAAGAAATCCAAACAGATAGAGCAAGTGGCTTTAGGGC 1524
QY 1778 CCATCTCTTCCCTCGGCTGATTTATCCCAAGAAATAGGATGTCCCAAGCAACACTTC 1837
Db 1525 CCAGATCCCTTCCCTCAGGCTGTTTACCCAGGGAATAGGATGTCTGGGCAAGTTTCC 1584
QY 1838 CCAGCCAATCGAGTCTGATAGTCCAGTTATCAGAAAGATATGCTGTAAAGTGTGATG 1897
Db 1585 C---CTAAGTGAAGTGTGATAAGTCTGCTTATCAGAAAGATATTACTGGGGGTGTGATA 1641
QY 1898 CACAGTGC--TTGCATTTTCTTGATAGCTTAGTCAATATAGAGCTGACAAAGAGAAAA 1955
Db 1642 TGTAGGGCATCTACATTTTCTTGATA--GGTAGTCATATGAAGCTGACAAAGAA--AAAA 1698
QY 1956 AGAGCAGCGATGTGGTGCAATATTAACAGCAGCTGTCTCCCTGGCTTCCGATAGCTGGG 2015
Db 1699 AGGCAGTGTGATGTGGCAATGTCAACAGACAGCTGTCCCTGTAC--TCTTGACAATAGG 1757
QY 2016 ATGACTCGCAATTCGAGCGGTGTGGTCTACTGCCAAAGGAATGACCTCTCACTTCTT 2075
Db 1758 ATGACTTGCATTCGCTGAGCGATGTGATCACCACCAAGGAATGGCCCTCTCACATTTCT 1817
QY 2076 CTTGATTCGCATACGCCGGG-----CCAGCTTGTCTACTCTCCCTTGGCTTCCAGAC 2130
Db 1818 CTTGATTCACATTTTCAGCAGGGTGTAGCTTGTCTCCCTCTCCCTCTTCACTTCCAGAC 1877
QY 2131 ACTAAGTCTGGAATGAAAAATTCACCTGCTCTGAATTTGGCCACTGGTGGGGCAGGGGTG 2190

Db 1878 ACTGAGTCTGGAATGAAAAATTCACCTGCTCTGAGTTGGCTCTAATGGGGCGGAGTG 1937
QY 2191 TGACTTTGGCTTCCAGGCTTGGAGGATTATCTCACCAGCCCTAGCTATATATA--CGGGCTG 2249
Db 1938 TTACTTTGGTTTCCAGGTTTGGAGGATTATCTCACCAGCCCTAGCTATATAAGCTGACCG 1997
QY 2250 GTGTGGAGGGGCTCCACAGGGCCAGTTCCAGGGGTTTCATCCACAGAGAGAAAAACATAG 2309
Db 1998 GTGTGGAGGGGCTCCACAGGGCCAACTCCAGGATTCTCTC--CAGCAGAAAAAACATAC 2056
QY 2310 A 2310
Db 2057 A 2057

RESULT 3
US-10-175-523-80
; Sequence 80, Application US/10175523
; Publication No. US20030096264A1
; GENERAL INFORMATION:
; APPLICANT: Brockman, Jeffrey
; APPLICANT: Evans, David
; APPLICANT: Hook, Derek
; APPLICANT: Klimczak, Leszek
; APPLICANT: Laeng, Pascal
; APPLICANT: Palfreyman, Michael
; APPLICANT: Rajan, Prithi
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
; FILE REFERENCE: 3235/J0795-US3
; CURRENT APPLICATION NUMBER: US/10/175,523
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: US 60/299,151
; PRIOR FILING DATE: 2001-06-18
; PRIOR APPLICATION NUMBER: US 60/317,828
; PRIOR FILING DATE: 2001-09-07
; PRIOR APPLICATION NUMBER: US 60/325,150
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 60/333,047
; PRIOR FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: US 60/349,936
; PRIOR FILING DATE: 2002-01-18
; PRIOR APPLICATION NUMBER: US 60/361,834
; PRIOR FILING DATE: 2002-03-04
; NUMBER OF SEQ ID NOS: 197
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 80
; LENGTH: 90650
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-175-523-80

Query Match 4.3%; Score 100.4; DB 14; Length 90650;
Best Local Similarity 74.7%; Pred. No. 5.7e-18;
Matches 139; Conservative 0; Mismatches 46; Indels 1; Gaps 1;

QY 198 TCTAGGCGCAGAGTCTCAACCTGGGGCCCTCGACCCCTTTGGGGGAATCAACACCCCT 257
Db 56512 TCTATAGCAGTGGTCTCAACCGAGTGGCTCAACCCCTATGTTGGTGG-CTGAATGACCTT 56570
QY 258 TTACAGGGGTCAACATATCATCTATCTCTATATGTCAAGGTATTACATTACGATTTCGTAACA 317
Db 56571 TCATAGAGTCAACATATCTGATATCTTGCATATCATATTTACATTAGGATTCAATAACA 56630
QY 318 GTAGCAAAATTACAGGTATGAAATAGCAATGAAATTAATTTTATGATTGAAGTCAACACA 377
Db 56631 GTAGGAAAGTTACAGTCAATGAGTAGCAATGAAATAATTTTATGTTTTTGGTGACCACA 56690
QY 378 ACATGA 383
Db 56691 TGAGGA 56696
```

Qy	324	AAATTACAGGTATGAATAAGCAATGAAATAATTTTA-----TCATTGAAGGTCCACCACAAC	379
Db	208947	AAACTACAGCTATGAAGTACCAATGAAATAATCTTACAGTTGGGGGAGGTCCACCACGC	209006
Qy	380	ATGAGG 385	
Db	209007	ATGAGG 209012	
RESULT 5			
US-10-175-523-97			
; Sequence 97, Application US/10175523			
; Publication No. US20030096264A1			
; GENERAL INFORMATION:			
; APPLICANT: Brockman, Jeffrey			
; APPLICANT: Evans, David			
; APPLICANT: Hook, Derek			
; APPLICANT: Klimczak, Leszek			
; APPLICANT: Laeng, Pascal			
; APPLICANT: Palfreyman, Michael			
; APPLICANT: Rajan, Priithi			
; TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)			
; FILE REFERENCE: 3235/13795-US3			
; CURRENT APPLICATION NUMBER: US/10/175,523			
; CURRENT FILING DATE: 2002-06-18			
; PRIOR APPLICATION NUMBER: US 60/299,151			
; PRIOR FILING DATE: 2001-06-18			
; PRIOR APPLICATION NUMBER: US 60/317,828			
; PRIOR FILING DATE: 2001-09-07			
; PRIOR APPLICATION NUMBER: US 60/325,150			
; PRIOR FILING DATE: 2001-09-25			
; PRIOR APPLICATION NUMBER: US 60/333,047			
; PRIOR FILING DATE: 2001-11-14			
; PRIOR APPLICATION NUMBER: US 60/349,936			
; PRIOR FILING DATE: 2002-01-18			
; PRIOR APPLICATION NUMBER: US 60/361,834			
; PRIOR FILING DATE: 2002-03-04			
; NUMBER OF SEQ ID NOS: 197			
; SOFTWARE: PatentIn version 3.1			
; SEQ ID NO 97			
; LENGTH: 106664			
; TYPE: DNA			
; ORGANISM: Mus musculus domesticus			
US-10-175-523-97			
Query Match 3.6%; Score 84.2; DB 14; Length 106664;			
Best Local Similarity 68.8%; Pred. No. 6e-13; Mismatches 1; Gaps 1;			
Matches 130; Conservative 0;			
Qy	198	TCTAGGGCAGCAGTCTCTCAACCTGGGGGCGCTCGACCCCTTTGGGGGAATCAAACGACCCCT	257
Db	35071	TCTAATACAAATGATCTTAACTCTGTGGGAGTGCGCTCTTTAGGAGGTTCAAAGGACCCCT	35130
Qy	258	TTACAGGGGTCAATCATCATCTATCTATCTATCTAGGTATTTTACATTCAGATTCGTATAACA	317
Db	35131	TTCATAGGGGTTCATAGACCATCGAAAACATAGATATTATACATTTATGATTCATAATA	35190
Qy	318	GTAGCAAAATACAGGTATGAATAGCAAT-CAAATATTTTATCATGATTCAGGTCCACCAC	376
Db	35191	GTAGCAAAATTTACATTTATGAGTAGAACATAAAAAATAATTTTTATGTTGGGGGTCCACCAC	35250
Qy	377	AACATGAGG 385	
Db	35251	AACATGAGG 35259	
RESULT 6			
US-09-771-208-20/c			
; Sequence 20, Application US/09771208			
; Patent No. US20020155564A1			
; GENERAL INFORMATION:			
; APPLICANT: MEDRANO, JUAN			
; APPLICANT: BRADFORD, ERIC			
; APPLICANT: HORVAT, SIMON			
; TITLE OF INVENTION: CLONING OF A HIGH-GROWTH GENE			
; FILE REFERENCE: 407T-923710US			
; CURRENT APPLICATION NUMBER: US/09/771,208			
; CURRENT FILING DATE: 2001-01-26			
; PRIOR APPLICATION NUMBER: US 08/999,477			
; PRIOR FILING DATE: 1997-12-29			
; NUMBER OF SEQ ID NOS: 20			
; SOFTWARE: PatentIn version 3.0			
; SEQ ID NO 20			
; LENGTH: 659158			
; TYPE: DNA			
; ORGANISM: Mus musculus			
FEATURE:			
NAME/KEY: misc feature			
LOCATION: (123459)..(123478)			
OTHER INFORMATION: n is unidentified a, c, g, or t			
NAME/KEY: misc feature			
LOCATION: (602466)..(602485)			
OTHER INFORMATION: n is unidentified a, c, g, or t			
NAME/KEY: misc feature			
LOCATION: (546398)..(547017)			
OTHER INFORMATION: n is unidentified a, c, g, or t			
NAME/KEY: misc feature			
LOCATION: (494715)..(494814)			
OTHER INFORMATION: n is unidentified a, c, g, or t			
NAME/KEY: misc feature			
LOCATION: (390366)..(391005)			
OTHER INFORMATION: n is unidentified a, c, g, or t			
NAME/KEY: misc feature			
LOCATION: (346860)..(346823)			
OTHER INFORMATION: n is unidentified a, c, g, or t			
NAME/KEY: misc feature			
LOCATION: (317174)..(317193)			
OTHER INFORMATION: n is unidentified a, c, g, or t			
NAME/KEY: misc feature			
LOCATION: (280353)..(280373)			
OTHER INFORMATION: n is unidentified a, c, g, or t			
NAME/KEY: misc feature			
LOCATION: (271829)..(271848)			
OTHER INFORMATION: n is unidentified a, c, g, or t			
NAME/KEY: misc feature			
LOCATION: (183872)..(183891)			
OTHER INFORMATION: n is unidentified a, c, g, or t			
NAME/KEY: misc feature			
LOCATION: (170625)..(170645)			
OTHER INFORMATION: n is unidentified a, c, g, or t			
NAME/KEY: misc feature			
LOCATION: (132680)..(132700)			
OTHER INFORMATION: n is unidentified a, c, g, or t			
NAME/KEY: misc feature			
OTHER INFORMATION: n is a, c, g, or t			
US-09-771-208-20			
Query Match 3.8%; Score 88.8; DB 10; Length 659158;			
Best Local Similarity 74.7%; Pred. No. 8.5e-14;			
Matches 139; Conservative 0; Mismatches 42; Indels 5; Gaps 2;			
Qy	204	GCAGCAGTCTCAACCTGGGGGCGCTCGACCCCTTTGGGGGAATCAAACGACCCCTTTACAG	263
Db	208828	GCAGCAGTCTCCAACCTGGGGGCGCTCGACATCTCTTTGGGAG-CCGAAAGGGGCTTTTCACAG	208886
Qy	264	GGGTACATATCATCTATCTATATGTCAGGTATTTTACATTCAGTTCGTAAACAGTAGCA	323
Db	208887	GGGTACAGGTGAGTACCTCGAATGTTGTCATATTTATTAATTCATAGCAGCTGCA	208946

; APPLICANT: HORVAT, SIMON
 ; TITLE OF INVENTION: CLONING OF A HIGH-GROWTH GENE
 ; FILE REFERENCE: 4077-923710US
 ; CURRENT APPLICATION NUMBER: US/09/771,208
 ; PRIOR FILING DATE: 2001-01-26
 ; PRIOR APPLICATION NUMBER: US 08/999,477
 ; PRIOR FILING DATE: 1997-12-29
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: PatentIn version 3.0.
 ; SEQ ID NO 20
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (123459)..(123478)
 ; OTHER INFORMATION: n is unidentified a, c, g, or t
 ; NAME/KEY: misc feature
 ; LOCATION: (602466)..(602485)
 ; OTHER INFORMATION: n is unidentified a, c, g, or t
 ; NAME/KEY: misc feature
 ; LOCATION: (546598)..(547017)
 ; OTHER INFORMATION: n is unidentified a, c, g, or t
 ; NAME/KEY: misc feature
 ; LOCATION: (494715)..(494814)
 ; OTHER INFORMATION: n is unidentified a, c, g, or t
 ; NAME/KEY: misc feature
 ; LOCATION: (390586)..(391005)
 ; OTHER INFORMATION: n is unidentified a, c, g, or t
 ; NAME/KEY: misc feature
 ; LOCATION: (346860)..(346823)
 ; OTHER INFORMATION: n is unidentified a, c, g, or t
 ; NAME/KEY: misc feature
 ; LOCATION: (317174)..(317193)
 ; OTHER INFORMATION: n is unidentified a, c, g, or t
 ; NAME/KEY: misc feature
 ; LOCATION: (280353)..(280373)
 ; OTHER INFORMATION: n is unidentified a, c, g, or t
 ; NAME/KEY: misc feature
 ; LOCATION: (271829)..(271848)
 ; OTHER INFORMATION: n is unidentified a, c, g, or t
 ; NAME/KEY: misc feature
 ; LOCATION: (183872)..(183891)
 ; OTHER INFORMATION: n is unidentified a, c, g, or t
 ; NAME/KEY: misc feature
 ; LOCATION: (170625)..(170645)
 ; OTHER INFORMATION: n is unidentified a, c, g, or t
 ; NAME/KEY: misc feature
 ; LOCATION: (132680)..(132700)
 ; OTHER INFORMATION: n is unidentified a, c, g, or t
 ; NAME/KEY: misc feature
 ; OTHER INFORMATION: n is a, c, g, or t
 ; US-09-771-208-20

Query Match 3.5%; Score 82; DB 10; Length 659158;
 Best Local Similarity 71.8%; Pred. No. 1e-11;
 Matches 135; Conservative 0; Mismatches 50; Indels 3; Gaps 2;
 QY 198 TTAGGGGAGCAGTCTCAACCTGGGGGCTCGACCCCTTTGGGGAATCAACGACCCT 257
 Db 439418 TTAGACTAGTGGTTCTCAACCTATGGGTTTCAACCTCTTTGGGAGTTTCATATCAGATA 439359
 QY 258 TTACAGGGGTACATATCATCTATATGTCAGGTATTTACATGATTTCGTAACA 317
 Db 439358 TT--CTGCGTAGCATATCAGATCTCGATATCAGATATTTCCATATGACACATAATG 439301
 QY 318 GTAGCAAAATACAGGTATGAATAGCAATGAATAATTTATGATTGAAGGTCCACCA 377
 Db 439300 GTAGCAAAATACAGGTATGAGGTAACATGAATGATTTATGTTG-GGGTCACCGCA 439242
 QY 378 ACATGAGG 385
 Db 439241 ACATGTGG 439234

RESULT 7

US-10-002-631C-156
 ; Sequence 156, Application US/10002631C
 ; Publication No. US20030157486A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Graff, Jonathan M.
 ; TITLE OF INVENTION: METHODS TO IDENTIFY SIGNAL SEQUENCES
 ; FILE REFERENCE: A34943 090495.0243
 ; CURRENT APPLICATION NUMBER: US/10/002,631C
 ; PRIOR FILING DATE: 2001-10-31
 ; PRIOR APPLICATION NUMBER: 60/300,309
 ; PRIOR FILING DATE: 2001-06-21
 ; NUMBER OF SEQ ID NOS: 324
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 156
 ; LENGTH: 889
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; FEATURE:
 ; NAME/KEY: unsure
 ; LOCATION: (1)...(203)
 ; OTHER INFORMATION: n = A, C, G or T
 ; US-10-002-631C-156

Query Match 3.4%; Score 80.6; DB 12; Length 889;
 Best Local Similarity 74.8%; Pred. No. 2.5e-13;
 Matches 101; Conservative 0; Mismatches 34; Indels 0; Gaps 0;
 QY 251 CGACCCCTTTACAGGGGTACATATCATCTATATGTCAGGTATTTACATGATT 310
 Db 441 CAACACCCCTTTGGGGTCTATATCGATATCTGCATATCCAATATTTACATGACATT 500
 QY 311 CGTAACAGTAGCAAAATTTACAGGTATGAATAGCAATGAATAATTTATGATTGAAGGT 370
 Db 501 CACAACAGCGCAAAATTTACAGGTATGAAGTAGCAACAAATAAATTTAGGGTTGGGAT 560
 QY 371 CACCACACATGAGG 385
 Db 561 CACCACGACATGAGG-575

RESULT 8

US-10-195-963-1
 ; Sequence 1, Application US/10195963
 ; Publication No. US20030167488A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Roberts, Steven L
 ; APPLICANT: Huff, Rita M
 ; TITLE OF INVENTION: MICE HETEROZYGOUS FOR WFS1 GENE AS MOUSE MODELS FOR
 ; TITLE OF INVENTION: DEPRESSION
 ; FILE REFERENCE: 28341/6284.N
 ; CURRENT APPLICATION NUMBER: US/10/195,963
 ; CURRENT FILING DATE: 2002-07-16
 ; PRIOR APPLICATION NUMBER: US/09/871,107
 ; PRIOR FILING DATE: 2001-05-31
 ; NUMBER OF SEQ ID NOS: 13
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 10917
 ; TYPE: DNA
 ; ORGANISM: Mus musculus
 ; FEATURE:
 ; NAME/KEY:
 ; LOCATION:
 ; OTHER INFORMATION: m=a or c; w=a or t; y=c or t; n=g, a, c or t
 ; US-10-195-963-1

Query Match 3.4%; Score 80.4; DB 12; Length 10917;
 Best Local Similarity 70.8%; Pred. No. 1.7e-12;
 Matches 136; Conservative 0; Mismatches 51; Indels 5; Gaps 2;

Qy	197	TTCTAGGCAGCAGTTCTCAACCTGGGGCCTCGACCCCTTTGGGGGAATCAAAACGACCC	256
Db	10340	TGCGAGGACAGAGGTTCTCAACCTGTGGGGCACAACCCCTTTGGGG--TTTGAATGACTT	10397
Qy	257	TTTACAGGGGTGCATATACA---TCTATCTTATGTCAAGTATTTACATTACGATTGCT	313
Db	10398	TTCACAGGGGTTGCTTGCCTCCAAAGACCATCAGAAAACACAGGTATTTACATTCCTCAATTCAT	10457
Qy	314	AACAGTAGCAAAATTACAGGTTATGAATAGCAATGAAATAATTTTATGATTTGAAGGTCAAC	373
Db	10458	GACAGTAGTAAATTTACAGTTATGAGTAGCAATGAAATAATTTTATGTTGGGTGAC	10517
Qy	374	CACACATGAGG	385
Db	10518	CACATCATGAAG	10529

RESULT 9
US-10-175-523-62
Sequence 62, Application US/10175523
Publication No. US2003096264A1
GENERAL INFORMATION:
APPLICANT: Brockman, Jeffrey
APPLICANT: Evans, David
APPLICANT: Hook, Derek
APPLICANT: Klimczak, Leszek
APPLICANT: Laeng, Pascal
APPLICANT: Palfreyman, Michael
APPLICANT: Rajan, Prithi
TITLE OF INVENTION: MULTI-PARAMETER HIGH THROUGHPUT SCREENING ASSAYS (MPHTS)
FILE REFERENCE: 3235/1J795-US3
CURRENT APPLICATION NUMBER: US/10/175,523
CURRENT FILING DATE: 2002-06-18
PRIOR APPLICATION NUMBER: US 60/299,151
PRIOR FILING DATE: 2001-06-18
PRIOR APPLICATION NUMBER: US 60/317,828
PRIOR FILING DATE: 2001-09-07
PRIOR APPLICATION NUMBER: US 60/325,150
PRIOR FILING DATE: 2001-09-25
PRIOR APPLICATION NUMBER: US 60/333,047
PRIOR FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: US 60/349,936
PRIOR FILING DATE: 2002-01-18
PRIOR APPLICATION NUMBER: US 60/361,834
PRIOR FILING DATE: 2002-03-04
NUMBER OF SEQ ID NOS: 197
SOFTWARE: PatentIn version 3.1
SEQ ID NO 62
LENGTH: 185548
TYPE: DNA
ORGANISM: Mus musculus
US-10-175-523-62

Query Match	3.3%;	Score 78.2;	DB 14;	Length 185548;
Best Local Similarity	77.2%;	Pred. No. 6.2e-11;		
Matches	95;	Conservative 0;	Mismatches 28;	Indels 0; Gaps 0;

QY	263	GGGGTCACATATCATCTATCCTATATGTCAGGTATTTACATTTACGATTTCGTTAACAGTAGC	322
Db	55602	GAGGTCACACATCAGATATCCTCCATATCAAAATATTTACATTTATAATTCATAACACTAGC	55661
QY	323	AAAAATTCAGGTATGAAATAGCAATGAAATAATTTTATGATTGAAGGTCAACCACAAACATG	382
Db	55662	AAAAATTAAGTTACAAGTAGCAACAAATAATATTTTATGGTAAGGGCCATCACAAATATG	55721
QY	383	AGG 385	
Db	55722	AGG 55724	

RESULT 10
US-10-228-794-75

```

; Sequence 75, Application US/10228794
; Publication No. US20030027198A1
; GENERAL INFORMATION:
; APPLICANT: VANDERBILT UNIVERSITY
; 305 Kirkland Hall
; Nashville, TN 37240
; TITLE OF INVENTION: MAMMALIAN GENES INVOLVED IN VIRAL
; INFECTION
; NUMBER OF SEQUENCES: 83
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Needle & Rosenberg, P.C.
; STREET: 127 Peachtree Street, Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303-1811
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/228,794
; FILING DATE: 27-Aug-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/171,209
; FILING DATE: 08-Mar-1999
; APPLICATION NUMBER: PCT/US97/06067
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Selby, Elizabeth
; REGISTRATION NUMBER: 38,298
; REFERENCE/DOCKET NUMBER: 22000.0061/P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 404 688 0770
; TELEFAX: 404 688 9880
; INFORMATION FOR SEQ ID NO: 75:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 892 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 75:
US-10-228-794-75

```

Query Match	3.3%;	Score 78;	DB 14;	Length 892;
Best Local Similarity	70.5%;	Pred. No. 1.5e-12;		
Matches 148;	Conservative 0;	Mismatches 55;	Indels 7;	Gaps 3;

Qy	177	CTCTTTTCAGGATCAGCCTGTATTCTAGGGCAGCAGTTCTCAACCTCGGGGGCTCGACCCCT	236
Db	119	CAGTACATAGATAGTCAAAATCTAGAGCACTGTTTCTATACCTGTGAGTTGCAACCCCT	178
Qy	237	TTGGG---GGAATCAAAAGCACCTT-TTACAGGGGTCACATATCATCTATCTATATGTCA	292
Db	179	TTGGGAGTGGGTCAAATGACCCCTATCACAGGGGTCTCAAATGAGATATCCTGCATATCA	238
Qy	293	GGTATTTCATATTACGATTTCGTAACTAGCAGCAAAATTCACAGGTATGAAATAGCAATGAAAT	352
Db	239	AATATTTCATATTGATTTCAATAGTAGTACCGAATTCACAGTTATGAAGTTACA---AAT	295
Qy	353	AATTTTATGATTGAAGTCAACCAACATG	382
Db	296	AATTTTATAGCTGAGAGTCACCAACAATG	325

RESULT 11
US-10-026-188-3/c
; Sequence 3, Application US/10026188
; Publication No. US20020164645A1
; GENERAL INFORMATION:

```
; APPLICANT: Zuker, Charles S.
; APPLICANT: Zhang, Yifeng
; APPLICANT: The Regents of the University of California
; TITLE OF INVENTION: Assays for Taste Receptor Cell Specific
; FILE OF INVENTION: Ion Channel
; FILE REFERENCE: 02307E-114910US
; CURRENT APPLICATION NUMBER: US/10/026,188
; CURRENT FILING DATE: 2001-12-21
; PRIOR FILING DATE: 2000-12-29
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 249487
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: mouse genomic region containing ltrpc5
US-10-026-188-3

Query Match          3.3%; Score 76.8; DB 13; Length 249487;
Best Local Similarity 70.2%; Pred. No. 2e-10;
Matches 146; Conservative 0; Mismatches 57; Indels 5; Gaps 3;

QY 179 CTTTCAGGATCAGCCTGATCTTAGGCGCAGCAGTCTTCAACCTGGGGGCGCTCGACCCCTTT 238
DB 238599 CCTACAGCCCTTCCTAGCCCTACACGAGTGGTCTTCAACCTGTGCTTTGAGACCCCTTT 238540

QY 239 GGGGGAATCAACAGCCCTTACAGGGGTACACATATCATCTATCTATATGTCAGGTATT 298
DB 238539 --GGAAGTCATGACCCCTTCAAGGGGTGCTTAAGACCATC--AGAAAAATAGATATT 238484

QY 299 TACATTACGATTCGTAAACAGTAGCAGCAAAATACAGGTATGAAATAGCAAT-GAAATAAATTT 357
DB 238483 TACATTATGATTCATAGCAGTAGCAGCAAAATATAGTTATGAGTAGCAATAAATAAATTT 238424

QY 358 TATGATTGAAGTCCACCAACAATGAGG 385
DB 238423 TGTGGTTGAGATCACCACAACATGAGG 238396

RESULT 12
US-09-728-445-91
; Sequence 91, Application US/09728445
; Patent No. US20020102543A1
; GENERAL INFORMATION:
; APPLICANT: Friedrich, Glenn
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; TITLE OF INVENTION: No. US20020102543A1el Mutated Mammalian Cells and
; FILE OF INVENTION: Animals
; FILE REFERENCE: LEX-0102-USA
; CURRENT APPLICATION NUMBER: US/09/728,445
; CURRENT FILING DATE: 2000-11-30
; PRIOR FILING DATE: 1999-12-01
; NUMBER OF SEQ ID NOS: 891
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 91
; LENGTH: 335
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)....(335)
; OTHER INFORMATION: n = A,T,C or G
US-09-728-445-91

Query Match          3.2%; Score 75.8; DB 10; Length 335;
Best Local Similarity 81.8%; Pred. No. 3.6e-12;
Matches 112; Conservative 0; Mismatches 22; Indels 3; Gaps 2;

QY 210 GTTCTCAACCTGGGGGCGCTCGACCCCTTTGGGGGAATCAACGAGCCCTTT-ACAGGGGTC 268
```

```
DB 201 GTTCTCAACCTGTGGGTACAAACCCCTTTGAGGG--TCAAACGACCCCTTTTCAAGAGTT 258
QY 269 ACATATCATCTATCTTATATGTCAGGTATTTACATTACGATTCGTAAACAGTAGCAAAATTT 328
DB 259 AGTATTAGCTATCTCGGATGTCAGATATTTTACATTCAATGTATTAACAGCAGCAAAATTT 318
QY 329 ACAGGTATGAATAAGCA 345
DB 319 ACAGTTATGAAGTAGCA 335

RESULT 13
US-09-917-800A-477/c
; Sequence 477, Application US/09917800A
; Patent No. US20020119462A1
; GENERAL INFORMATION:
; APPLICANT: Mendrick, Donna
; APPLICANT: Porter, Mark
; APPLICANT: Johnson, Kory
; APPLICANT: Castle, Arthur
; APPLICANT: Blashoff, Michael
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Molecular Toxicology Modeling
; FILE REFERENCE: 44921-5038-US
; CURRENT APPLICATION NUMBER: US/09/917,800A
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/222,040
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: US 60/222,880
; PRIOR FILING DATE: 2000-11-02
; PRIOR APPLICATION NUMBER: US 60/290,029
; PRIOR FILING DATE: 2001-05-11
; PRIOR APPLICATION NUMBER: US 60/290,645
; PRIOR FILING DATE: 2001-05-15
; PRIOR APPLICATION NUMBER: US 60/292,336
; PRIOR FILING DATE: 2001-05-22
; PRIOR APPLICATION NUMBER: US 60/295,798
; PRIOR FILING DATE: 2001-06-06
; PRIOR APPLICATION NUMBER: US 60/297,457
; PRIOR FILING DATE: 2001-06-13
; PRIOR APPLICATION NUMBER: US 60/298,884
; PRIOR FILING DATE: 2001-06-19
; PRIOR APPLICATION NUMBER: US 60/303,459
; NUMBER OF SEQ ID NOS: 1740
; SOFTWARE: PatencIn Ver. 2.1
; SEQ ID NO 477
; LENGTH: 5990
; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020119462A1 AB009636
US-09-917-800A-477

Query Match          3.2%; Score 75.6; DB 10; Length 5990;
Best Local Similarity 66.7%; Pred. No. 3.3e-11;
Matches 124; Conservative 0; Mismatches 59; Indels 3; Gaps 1;

QY 196 ATTCTAGGCGACGATTTCTCAACCTGGGGGCGCTCGACCCCTTTGGGGGAATCAACGACC 255
DB 4978 ACTCTAAGCCTGTGGTTCTCAACCCCATGAGTCATGACCCCTTTGGGATTGTCCTCAATGACCC 4919

QY 256 CTTT---ACAGGGGTACATATCATCTATCTATATGTCAGGTATTTACATTACGATTCG 312
DB 4918 TTTTATCAGGGGTGCAATAGTAGTATCTCTGTAGATCAGATATATACATTGCAATTTCA 4859

QY 313 TAACAGTAGCAAAATACAGGTATGAAATAGCAATGCAAAATATTTTATGTAAGGTCA 372
DB 4858 TAACCGTGTCAAAATACAGTTGTGGAATAGCAACTAAATAATTTTGTGGTTTGGGTCA 4799
QY 373 CCACAA 378
```

Db 4798 CTACAA 4793

RESULT 14
US-09-901-484A-183
Sequence 183, Application US/09901484A
Patent No. US20020119460A1
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Chumakov, Ilya
APPLICANT: Bougueleret, Lydie
TITLE OF INVENTION: Prostate Cancer Gene
FILE REFERENCE: GEN-T11XC3D2
CURRENT APPLICATION NUMBER: US/09/901,484A
CURRENT FILING DATE: 2001-07-09
PRIOR APPLICATION NUMBER: US 08/996,306
PRIOR FILING DATE: 1997-12-22
PRIOR APPLICATION NUMBER: US 60/099,658
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: US 09/218,207
PRIOR FILING DATE: 1998-12-22
PRIOR APPLICATION NUMBER: US 09/338,907
PRIOR FILING DATE: 1999-06-23
PRIOR APPLICATION NUMBER: US 09/853,526
PRIOR FILING DATE: 2001-05-11
NUMBER OF SEQ ID NOS: 578
SOFTWARE: PatentIn version 3.1
SEQ ID NO 183
LENGTH: 37950
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: misc feature
LOCATION: (616)..(616)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature
LOCATION: (1552)..(1552)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature
LOCATION: (2809)..(2809)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature
LOCATION: (2821)..(2824)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature
LOCATION: (2826)..(2826)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature
LOCATION: (2828)..(2829)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature
LOCATION: (2831)..(2833)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature
LOCATION: (2835)..(2835)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature
LOCATION: (5259)..(5328)
OTHER INFORMATION: exon 2
NAME/KEY: misc feature
LOCATION: (6247)..(6247)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature
LOCATION: (8667)..(8667)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature
LOCATION: (8669)..(8669)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature
LOCATION: (8671)..(8679)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature

LOCATIONS: (8681)..(8685)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature
LOCATION: (8687)..(8687)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature
LOCATION: (8689)..(8693)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature
LOCATION: (8698)..(8698)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature
LOCATION: (9062)..(9062)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature
LOCATION: (12298)..(12298)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature
LOCATION: (12675)..(12791)
OTHER INFORMATION: exon 3
NAME/KEY: misc feature
LOCATION: (14621)..(14710)
OTHER INFORMATION: exon 4
NAME/KEY: misc feature
LOCATION: (17578)..(17578)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature
LOCATION: (17639)..(17639)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature
LOCATION: (17707)..(17707)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature
LOCATION: (17867)..(17867)
OTHER INFORMATION: n = a, c, g, or t.
NAME/KEY: misc feature
LOCATION: (19822)..(19912)
OTHER INFORMATION: exon 5
NAME/KEY: misc feature
LOCATION: (21789)..(21950)
OTHER INFORMATION: exon 6
NAME/KEY: misc feature
LOCATION: (23387)..(23510)
OTHER INFORMATION: exon 7
NAME/KEY: misc feature
LOCATION: (25520)..(26016)
OTHER INFORMATION: exon 8
NAME/KEY: misc feature
LOCATION: (37931)..(37931)
OTHER INFORMATION: n = a, c, g, or t.

US-09-901-484A-183

Query Match 3.2%; Score 74.8; DB 10; Length 37950;
Best Local Similarity 80.0%; Pred. No. 2.2e-10;
Matches 88; Conservative 0; Mismatches 22; Indels 0; Gaps 0;

QY 276 ATCTATCCTATGTCAGGTATTACATTGCAGTATTTACATTGGTAACAGTAGCAAAAATTACAGGA 335
DB 9570 ACCTATCTCTTATCATAGTAGTTACATTGATTTAACAGGAGGACAAATCACAGTTA 9629

QY 336 TGAATAGCAATGAAATAATTTTTATGATTGAAGGTCACCACAATCAGG 385
DB 9630 CGCAATATCAACAAAAATATTTTATGTTTGAGGTCACCATAACGTCAGG 9679

RESULT 15
US-09-853-526-183
Sequence 183, Application US/09853526
Patent No. US20020165345A1
GENERAL INFORMATION:
APPLICANT: Cohen, Daniel
APPLICANT: Blumenfeld, Marta
APPLICANT: Ilyia, Chumakov

APPLICANT: Bougueleret, Lydie
TITLE OF INVENTION: PROSTATE CANCER GENE
FILE REFERENCE: GENSET 18CP1CP
CURRENT APPLICATION NUMBER: US/09/853,526
CURRENT FILING DATE: 2001-05-11
PRIOR APPLICATION NUMBER: 09/338,907
PRIOR FILING DATE: 1999-06-23
PRIOR APPLICATION NUMBER: 08/996,306
PRIOR FILING DATE: 1997-12-22
PRIOR APPLICATION NUMBER: 60/099,658
PRIOR FILING DATE: 1998-09-09
PRIOR APPLICATION NUMBER: 09/218,207
PRIOR FILING DATE: 1998-12-22
NUMBER OF SEQ ID NOS: 578
SOFTWARE: Patent.pm
SEQ ID NO 183

LENGTH: 37950

TYPE: DNA

ORGANISM: Mus musculus

FEATURE:

NAME/KEY: exon

LOCATION: 5259..5328

OTHER INFORMATION: exon2

NAME/KEY: exon

LOCATION: 12675..12791

OTHER INFORMATION: exon3

NAME/KEY: exon

LOCATION: 14621..14710

OTHER INFORMATION: exon4

NAME/KEY: exon

LOCATION: 19822..19912

OTHER INFORMATION: exon5

NAME/KEY: exon

LOCATION: 21789..21950

OTHER INFORMATION: exon6

NAME/KEY: exon

LOCATION: 23387..23510

OTHER INFORMATION: exon7

NAME/KEY: exon

LOCATION: 25520..26016

OTHER INFORMATION: exon8

US-09-853-526-183

Query Match 3.2%; Score 74.8; DB 10; Length 37950;
Best Local Similarity 80.0%; Pred. No. 2.2e-10;
Matches 88; Conservative 0; Mismatches 22; Indels 0; Gaps 0;
Qy 276 ATCTATCCTATATGTCAGGTATTTACATTACGATTCGTAAACAGTAGCAAAATTCAGGTA 335
Db 9570 ACCTATCCTGCTTATCAGATAGTTACATTATGAATTGTAACAGCAGCAAAATTCAGGTA 9629
Qy 336 TGAATAGCAATGAATAATTTATGATTGAAGGTCCACCAACATGAGG 385
Db 9630 CGCAATATCAACAAATAATTTATGGTTGAGGTCACCATAACGTGAGG 9679

Search completed: November 18, 2003, 07:42:01
Job time : 1143.46 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: November 17, 2003, 23:31:10 ; Search time 4918.71 Seconds
(without alignments)
11651.427 Million cell updates/sec

Title: US-10-005-337A-1

Perfect score: 2358

Sequence: 1 ggcattcttcattttaaca.....caggtcgaggccaccatgg 2358

Scoring table: IDENTITY NUC

Gapop 10_0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*

2: em_esthum:*

3: em_estin:*

4: em_estmu:*

5: em_estov:*

6: em_estpl:*

7: em_estro:*

8: em_htc:*

9: gb_est1:*

10: gb_est2:*

11: gb_htc:*

12: gb_est3:*

13: gb_est4:*

14: gb_est5:*

15: em_estfun:*

16: em_estom:*

17: em_gss_hum:*

18: em_gss_inv:*

19: em_gss_pln:*

20: em_gss_vrt:*

21: em_gss_fun:*

22: em_gss_nam:*

23: em_gss_mus:*

24: em_gss_pro:*

25: em_gss_rod:*

26: em_gss_phg:*

27: em_gss_vtl:*

28: gb_gse1:*

29: gb_gse2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	595	24.8	600	28	BH043261
2	505.4	21.4	586	28	AZ295534
3	306.8	13.0	371	13	BY091041
4	269.6	11.4	353	13	BY079997

C	5	191.6	8.1	478	12	B1293043	BI293043 UI-R-DKO-
C	6	183	7.3	434	10	BF543101	BF543101 UI-R-AGI-
C	7	171	7.3	488	9	AW252440	AW252440 UI-R-BJO-
C	8	171	7.3	511	9	AW251306	AW251306 UI-R-BJO-
C	9	167.8	7.1	370	9	AI710529	AI710529 UI-R-AGI-
C	10	131.8	5.6	735	9	AU139209	AU139209 UI-R-AGI-
C	11	119.6	5.1	509	10	BE627514	BE627514 uu52e08.Y
C	12	116.2	4.9	538	28	AZ804735	AZ804735 2M0065113
C	13	116	4.9	522	28	AQ480395	AQ480395 RPCI-11-2
C	14	114.6	4.9	518	28	AQ114168	AQ114168 RPCI-23-4
C	15	113	4.8	596	28	AZ290439	AZ290439 RPCI-23-5
C	16	112.4	4.8	623	10	BA470798	BA470798 BB470798
C	17	108	4.6	577	28	AZ734079	AZ734079 RPCI-24-1
C	18	107.8	4.6	556	28	AZ373614	AZ373614 1M0125N21
C	19	107.8	4.6	589	28	BH360537	BH360537 CH230-164
C	20	107.2	4.5	839	28	BZ110003	BZ110003 CH230-164
C	21	107	4.5	369	28	AZ600690	AZ600690 1M0418D12
C	22	106.2	4.5	4565	11	AK052932	AK052932 Mus muscu
C	23	106	4.5	3534	11	AK038950	AK038950 Mus muscu
C	24	104	4.4	310	28	AZ939809	AZ939809 2M0198P19
C	25	103.6	4.4	585	14	CA882248	CA882248 K0996B07-
C	26	103.6	4.4	834	28	BZ185566	BZ185566 CH230-340
C	27	103.2	4.4	519	28	AZ488841	AZ488841 1M0319013
C	28	103.2	4.4	640	29	BZ239325	BZ239325 CH230-282
C	29	103	4.4	574	28	AZ563281	AZ563281 RPCI-23-2
C	30	102.6	4.4	422	10	BF412836	BF412836 UI-R-BT1-
C	31	102.6	4.4	810	29	BZ264838	BZ264838 CH230-315
C	32	102	4.3	764	28	AZ348698	AZ348698 1M0085D24
C	33	101.8	4.3	374	28	BZ189630	BZ189630 CH230-525
C	34	101.6	4.3	676	28	AZ496223	AZ496223 1M0332H21
C	35	101.2	4.3	594	28	AZ389177	AZ389177 1M0149M20
C	36	101	4.3	477	28	AZ778894	AZ778894 2M0014H04
C	37	101	4.3	596	28	AZ856151	AZ856151 2M0160A02
C	38	100.8	4.3	407	13	BY549634	BY549634
C	39	100.6	4.3	600	12	EG803092	EG803092 0194-80 M
C	40	100.4	4.3	591	28	AZ445982	AZ445982 1M0242B09
C	41	100.4	4.3	787	28	BZ128535	BZ128535 CH230-298
C	42	100.4	4.3	803	29	BZ234545	BZ234545 CH230-427
C	43	100.4	4.3	2591	11	AK036783	AK036783 Mus muscu
C	44	100.2	4.2	1085	29	BZ223957	BZ223957 CH230-446
C	45	100	4.2	473	28	AZ060257	AZ060257 RPCI-23-4

ALIGNMENTS

RESULT 1
BH043261
LOCUS
DEFINITION
RPCI-24-330A6.TJ RPCI-24 Mus musculus genomic clone RPCI-24-330A6,
genomic survey sequence.
ACCESSION
BH043261
VERSION
BH043261.1
KEYWORDS
GSS.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
REFERENCE
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 600)
AUTHORS
Zhao,S., Nierman,W., Malek,J., Shatsman,S., Akinret,B., Levins,M.,
Tsegaye,G., Geer,K., Krol,M., Shvartsbeyn,A., Gebregeorgis,E.,
Russell,D., de Jong,P. and Fraser,C.M.
TITLE
Mouse BAC End Sequences from Library RPCI-24
JOURNAL
Unpublished
COMMENT
Other_GSSs: RPCI-24-330A6.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@igr.org
Clones are derived from the mouse BAC library RPCI-24. For BAC

library availability, please contact Pieter de Jong
(pdejong@mail.cho.org). Clones may be purchased from BACPAC
Resources (<http://www.choi.org/bacpac/orderingframe.htm>). BAC end
page: http://www.tigr.org/tadb/bac_ends/mouse/bac_end_intro.html
Plate: 330 row: A column: 6
Seq primer: SP6
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..600
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-24-330A6"
/sex="Male"
/cell_type="Spleen/Brain"
/clone_lib="RPCI-24"
/notes="Vector: pTARBAC1; Site 1: BamH1; Site 2: BamH1;
RPCI-24 Mouse BAC Library produced by Pieter de Jong. The
library was cloned in the pTARBAC1 cloning vector at the
BamH1 sites using MboI partially digested male C57BL/6J
DNA."

BASE COUNT 158 a 139 c 142 g 161 t
ORIGIN
Query Match 24.8%; Score 585; DB 28; Length 600;
Best Local Similarity 100.0%; Pred. No. 1.4e-148;
Matches 585; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

211 TTCTCAACCTGGGGCTCGACCCCTTTGGGGAATCAACAGACCCTTTACAGGGTCAC 270
16 TTCTCAACCTGGGGCTCGACCCCTTTGGGGAATCAACAGACCCTTTACAGGGTCAC 75
271 ATATCATCTATCTATGTAGCTATTTATACAGTTCGTAACAGTAGCAAAATTAC 330
76 ATATCATCTATCTATGTAGCTATTTATACAGTTCGTAACAGTAGCAAAATTAC 135
331 AGTATGAATAGCAATGAATAATTTATGATTGAAGTTCACCAACATGAGGCGCC 390
136 AGTATGAATAGCAATGAATAATTTATGATTGAAGTTCACCAACATGAGGCGCC 195
391 ACATGTTCTAGAGAAATACCTCGGTGGGAAAGTTTGGAAAGCCTTTCTGTCCA 450
196 ACATGTTCTAGAGAAATACCTCGGTGGGAAAGTTTGGAAAGCCTTTCTGTCCA 255
451 TTCTTCAATTCACAAAGTGATGTTTACAGAAAGCCTTTACGCTGTTCTGTGGGCTC 510
256 TTCTTCAATTCACAAAGTGATGTTTACAGAAAGCCTTTACGCTGTTCTGTGGGCTC 315
511 TTAGTAGTCTAGTAGGAACCTGTATGTACAGGCTGCTTCTTATGGGTGGAGCCAAG 570
316 TTAGTAGTCTAGTAGGAACCTGTATGTACAGGCTGCTTCTTATGGGTGGAGCCAAG 375
571 CGATCGTGGGTGAGCGAAGCGCAACCTCACCTTCTAGCTCTGCATCATAGCAAGTA 630
376 CGATCGTGGGTGAGCGAAGCGCAACCTCACCTTCTAGCTCTGCATCATAGCAAGTA 435
631 GCCTAATGTTTCTGTGTAGTGTGTCATCTCTGTGAATCGAGATCCTTGGCCCTTGTGA 690
436 GCCTAATGTTTCTGTGTAGTGTGTCATCTCTGTGAATCGAGATCCTTGGCCCTTGTGA 495
691 ATTAGGAGGCCAAAATCTAGAGATTCAGAACTGCTCAGAGCCCGAGAGTCTTCTT 750
496 ATTAGGAGGCCAAAATCTAGAGATTCAGAACTGCTCAGAGCCCGAGAGTCTTCTT 555
751 CAAAGGAAAGGTCTCAACTCTAGCCCCCTTAGCTCTGAGTCAG 795
556 CAAAGGAAAGGTCTCAACTCTAGCCCCCTTAGCTCTGAGTCAG 600

RESULT 2
A2295534
LOCUS
A2295534 586 bp DNA linear GSS 27-JUL-2000

DEFINITION
RPCI-23-105P3.TJ RPCI-23 Mus musculus genomic clone RPCI-23-105P3,
genomic survey sequence.
ACCESSION
A2295534
VERSION
A2295534.1 GI:9537319
KEYWORDS
GSS.
SOURCE
Mus musculus (house mouse)
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 586)
Zhou, S., Nierman, W., Feidblyum, T., Malek, J., Shatsman, S., Akinret
B., Levins, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C.M.
Mouse BAC End Sequences from Library RPCI-23
Unpublished
Other GSSs: RPCI-23-105P3.TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieterdejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tadb/bac_ends/mouse/bac_end_intro.html
Plate: 105 row: P column: 3
Seq primer: SP6
Class: BAC ends.

FEATURES
source Location/Qualifiers
1..586
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-105P3"
/sex="Female"
/lab_host="DH10B"
/clone_lib="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBACE3.6; Site 1:
ECORI; Site 2: ECORI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of ECORI and ECORI Methyase. Size
selected DNA was cloned into the pBACE3.6 vector at the
ECORI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."

BASE COUNT 141 a 148 c 154 g 143 t
ORIGIN
Query Match 21.4%; Score 505.4; DB 28; Length 586;
Best Local Similarity 98.5%; Pred. No. 8.6e-127;
Matches 531; Conservative 0; Mismatches 6; Indels 2; Gaps 2;
1765 GTGGCACCAGACCCATCTCTTCCCTCGGCTGATTATCCCAAGATAGATGCC 1824
1 GTGGCACCAGACCCATCTCTTCCCTCGGCTGATTATCCCAAGATAGATGCC 59
1825 AAAGCAACTTCCAGCAACTGGAGTCTGATAGTCCAGTTATCAGAAAGATATGCC 1884
60 AAAGCAACTTCCAGCAACTGGAGTCTGATAGTCCAGTTATCAGAAAGATATGCC 119
1885 TGTAACTGTGATGCACAGTCTTGCATTTTCTGATACGTTAGTATGAGAGCTGACA 1944
120 TGTAACTGTGATGCACAGTCTTGCATTTTCTGATACGTTAGTATGATGACA 178
1945 AAGAAGAAAAAGAGCAGCGATGTGTGCAATATTAACAGGAGCTGTCCCTCGGCTTC 2004
179 AAGAAGAAAAAGAGCAGCGATGTGTGCAATATTAACAGGAGCTGTCCCTCGGCTTC 238
2005 CGATACGTGGGATGACTCGCAITTCGTGAGCGGTGTGTCCTCACTGCAAGGAATGACCC 2064

Db 239 CGATACCTGGGATGACTCGCATTCGTGAGCGGTGTGCTACTGCCAAAGGAATGACCCCTC 238
QY 2065 TCACATTTCTTCTCTGATTGCATACGCGCGGCGGAGCTTGTCTATCTCCCTCTTGGGCTTC 2124
Db 299 TCACATTTCTTCTCTGATTGCATACGCGCGGCGGAGCTTGTCTATCTCCCTCTTGGGCTTC 358
QY 2125 CCAGACACTAAGTCTGGAATGAAATTCACCTGCTCTGTAATGGCCACTGGTGGGGCA 2184
Db 359 CCAGACACTAAGTCTGGAATGAAATTCACCTGCTCTGTAATGGCCACTGGTGGGGCA 418
QY 2185 GGGGTGTGACTTGGCTTCCAGGCTGGAAGATTATCTCACCAGCCCTAGCTATATAACG 2244
Db 419 GGGGTGTGACTTGGCTTCCAGGCTGGAAGATTATCTCACCAGCCCTAGCTATATAACG 478
QY 2245 GGCTGGTGTGAGGGGCTCCACAGGCGGCTTCACGGGTTTCATCCACAGAGAGAAAA 2303
Db 479 GGCTGGTGTGAGGGGCTCCACAGGCGGCTTCACGGGTTTCATCCACAGAGAGAAAA 537

RESULT 3

BY091041
LOCUS
DEFINITION
Musculus cdna clone K630084E21 5', mRNA sequence.
BY091041
ACCESSION
VERSION
KEYWORDS
SOURCE
Mus musculus (house mouse)
Mus musculus

REFERENCE

Okazaki Y., Furuno M., Kasukawa T., Adachi J., Bono H., Kondo S.,
Nikaido I., Otsu N., Saito R., Suzuki H., Yamanaka I., Kiyosawa H.,
Yagi K., Tomaru Y., Hasegawa Y., Nogami A., Schonbach C.,
Gojobori T., Baldarelli R., Hill D.P., Bult C., Hume D.A.,
Quackenbush J., Schriml L.M., Kanapin A., Matsuda H., Batalov S.,
Beisel K.W., Blake J.A., Bradt D., Brusic V., Chothia C., Corbani
L.E., Cousins S., Dalla E., Dragani T.A., Fletcher C.F., Forrest
A., Frazer K.S., Gaasterland T., Gariboldi M., Gissi C., Godzik A.,
Gough J., Grimmond S., Gustincich S., Hirokawa N., Jackson I.J.,
Jarvis E.D., Kanai A., Kawaji H., Kawasawa Y., Kedzierski R.M.,
King B.L., Konagaya A., Kurochkin I.V., Lee Y., Lenhard B., Lyons
P.A., Maglott D.R., Maltais L., Marchionni L., McKenzie L., Miki
H., Nagashima T., Numata K., Okido T., Pavan W.J., Pertea G.,
Pesole G., Petrovsky N., Pillai R., Pontius J.U., Qi D.,
Ramachandran S., Ravasi T., Reed J.C., Reed D.J., Reid J., Ring
B.Z., Ringwald M., Sandelin A., Schneider C., Semple C.A., Setou
M., Shimada K., Sultana R., Takenaka Y., Taylor M.S., Teasdale
R.D., Tomita M., Verardo R., Wagner L., Wahlestedt C., Wang Y.,
Watanabe Y., Wells C., Wilming L.G., Wynshaw-Boris A., Yanagisawa
M., Yang I., Yang L., Yuan Z., Zavolan M., Zhu Y., Zimmer A.,
Carninci P., Hayatsu N., Hirozane-Kishikawa T., Konno H., Nakamura
M., Sakazume N., Sato K., Shiraki T., Waki K., Kawai J., Aizawa K.,
Arakawa T., Fukuda S., Hara A., Hashizume W., Imorani K., Ishii
Y., Itoh M., Kagawa I., Miyazaki A., Sakai K., Sasaki D., Shibata
K., Shinagawa A., Yasunishi A., Yoshino M., Waterston R., Lander
E.S., Rogers J., Birney E. and Hayashizaki Y.

AUTHORS

Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)

TITLE

22354683
MEDLINE
PUBMED
12466851

COMMENT

Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
URL: http://genome.gsc.riken.go.jp/

FEATURES

source

1. .371
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K630084E21"
/tissue_type="heart"
/dev_stage="10 days neonate"
/clone_lib="RIKEN full-length enriched, 10 days neonate
heart"

BASE COUNT 91 a 94 c 105 g 81 t
ORIGIN

Query Match

Best Local Similarity 13.0%; Score 306.8; DB 13; Length 371;

Matches 308; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2008 TAGCTGGGATGACTCGCATTCGTGAGCGGTGTGCTACTGCCAAAGGAATGACCCCTCTCA 2067

Db 1 TAGCTGGGATGACTCGCATTCGTGAGCGGTGTGCTACTGCCAAAGGAATGACCCCTCTCA 60

QY 2068 CATTTCTTCTCTGATTCGCATACGCCGCGGCGGAGCTTGTCTCTCTTGGGCTTCCCA 2127

Db 61 CATTTCTTCTCTGATTCGCATACGCCGCGGCGGAGCTTGTCTCTCTTGGGCTTCCCA 120

QY 2128 GACACTTAAGTCTGGATGAATTCACCTGCTCTGAATGGCCACTGGTGGGGGAGGG 2187

Db 121 GACACTTAAGTCTGGATGAATTCACCTGCTCTGAATGGCCACTGGTGGGGGAGGG 180

QY 2188 GTGTGACTTGGCTTCCAGCGCTGGAAGATTATCTACCCAGCCCTAGCTATATAACGGGC 2247

Db 181 GTGTGACTTGGCTTCCAGCGCTGGAAGATTATCTACCCAGCCCTAGCTATATAACGGGC 240

QY 2248 TGGTGTGGAGGGCTCCACAGGCGGCTTCACAGGGTTCATCCACAGAGAGAAAAACAT 2307

Db 241 TGGTGTGGAGGGCTCCACAGGCGGCTTCACAGGGTTCATCCACAGAGAGAAAAACAT 300

QY 2308 AGACTCGAGG 2317

Db 301 AGACTCACGG 310

RESULT 4

BY079997

LOCUS

DEFINITION

musculus cdna clone K630022E19 5', mRNA sequence.

ACCESSION

BY079997

LOCUS

DEFINITION

musculus cdna clone K630022E19 5', mRNA sequence.

ACCESSION

BY079997

LOCUS

DEFINITION

musculus cdna clone K630022E19 5', mRNA sequence.

ACCESSION

BY079997

LOCUS

Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane
T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A.,
Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, H., Sakai, K.,
Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami
M., Waki, K., Watahiki, A., Muramatsu, M. and Hayashizaki, Y. Direct
Submission

Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)

Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.

Please visit our web site (<http://genome.gsc.riken.go.jp>) for
further details.

Location/Qualifiers

1. .371
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K630084E21"
/tissue_type="heart"
/dev_stage="10 days neonate"
/clone_lib="RIKEN full-length enriched, 10 days neonate
heart"

BASE COUNT 91 a 94 c 105 g 81 t
ORIGIN

Query Match 13.0%; Score 306.8; DB 13; Length 371;

Best Local Similarity 99.4%; Pred. No. 1.8e-72;

Matches 308; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 2008 TAGCTGGGATGACTCGCATTCGTGAGCGGTGTGCTACTGCCAAAGGAATGACCCCTCTCA 2067

Db 1 TAGCTGGGATGACTCGCATTCGTGAGCGGTGTGCTACTGCCAAAGGAATGACCCCTCTCA 60

QY 2068 CATTTCTTCTCTGATTCGCATACGCCGCGGCGGAGCTTGTCTCTCTTGGGCTTCCCA 2127

Db 61 CATTTCTTCTCTGATTCGCATACGCCGCGGCGGAGCTTGTCTCTCTTGGGCTTCCCA 120

QY 2128 GACACTTAAGTCTGGATGAATTCACCTGCTCTGAATGGCCACTGGTGGGGGAGGG 2187

Db 121 GACACTTAAGTCTGGATGAATTCACCTGCTCTGAATGGCCACTGGTGGGGGAGGG 180

QY 2188 GTGTGACTTGGCTTCCAGCGCTGGAAGATTATCTACCCAGCCCTAGCTATATAACGGGC 2247

Db 181 GTGTGACTTGGCTTCCAGCGCTGGAAGATTATCTACCCAGCCCTAGCTATATAACGGGC 240

QY 2248 TGGTGTGGAGGGCTCCACAGGCGGCTTCACAGGGTTCATCCACAGAGAGAAAAACAT 2307

Db 241 TGGTGTGGAGGGCTCCACAGGCGGCTTCACAGGGTTCATCCACAGAGAGAAAAACAT 300

QY 2308 AGACTCGAGG 2317

Db 301 AGACTCACGG 310

RESULT 4

BY079997

LOCUS

DEFINITION

musculus cdna clone K630022E19 5', mRNA sequence.

ACCESSION

BY079997

LOCUS

DEFINITION

musculus cdna clone K630022E19 5', mRNA sequence.

ACCESSION

BY079997

LOCUS

DEFINITION

musculus cdna clone K630022E19 5', mRNA sequence.

ACCESSION

BY079997

LOCUS

```

VERSION BY079997.1 GI:26185447
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 353)
Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S.,
Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamanaka, I., Kiyosawa, H.,
Yegi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C.,
Gojobori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A.,
Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S.,
Beisel, K.W., Blake, J.A., Bratt, D., Brusic, V., Chothia, C., Corbani,
L.E., Cousins, S., Dalla, E., Dragani, T.A., Fletcher, C.F., Forrest,
A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, J.,
Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J.,
Jarvis, E.D., Kanai, A., Kawai, H., Kawasawa, Y., Kedzierski, R.M.,
King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lenhard, B., Lyons,
P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki,
H., Nagashima, T., Numata, K., Okido, T., Pavan, W.J., Pertea, G.,
Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D.,
Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring,
B.Z., Ringwald, M., Sandelin, A., Schneider, C., Sempile, C.A., Setou,
M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale,
R.D., Tomita, M., Vezardo, R., Wagner, L., Wahlestedt, C., Wang, Y.,
Watanabe, Y., Wells, C., Wilming, L.G., Wyszewski, A., Yanagisawa,
M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A.,
Carninci, P., Hayatsu, N., Hirozane-Kishikawa, T., Konno, H., Nakamura,
M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K.,
Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii,
Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata,
K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander,
E.S., Rogers, J., Birney, E. and Hayashizaki, Y.
Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
Nature 420, 563-573 (2002)
12466851
22354683
PUBMED
COMMENT Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center (GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsr.riken.go.jp,
URL: http://genome.gsc.riken.go.jp/
Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hirozane,
T., Imotani, K., Ishii, Y., Itoh, M., Kawai, J., Konno, H., Miyazaki, A.,
Murata, M., Nakamura, M., Nomura, K., Numazaki, R., Ohno, M., Sakai, K.,
Sakazume, N., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tegami,
M., Waki, K., Watahiki, A., Muraizumi, M. and Hayashizaki, Y. Direct
Submission
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
Location/Qualifiers
1. .353

/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K63002E19"
/tissue_type="heart"
/dev_stage="10 days neonate"
/clone_lib="RIKEN full-length enriched, 10 days neonate
heart"
BASE COUNT 82 a 88 c 102 g 79 t 2 others
ORIGIN
Query Match 11.4%; Score 269.6; DB 13; Length 353;
Best Local Similarity 95.5%; Pred. No. 2.8e-62;
Matches 298; Conservative 0; Mismatches 11; Indels 3; Gaps 2;
QY 2006 GATAGCTGGATGACTCGGATTCGCTGAGCGGTGTGCTCACTGCCAAGGAATGACCCCTCT 2065
DB 2 GATAGCTGGATGACTCGGATTCGCTGAGCGGTGTGCTCACTGCCAAGGAATGACCCCTCT 61
QY 2066 CACATTTCTCTGATTCGCATACCGCGCGGCAGCTTCTCATCTCCCTCTTGGGCTTCC 2125
DB 62 CACATTTCTCTGATTCGCATACCGCGCGGCAGCTTCTCATCTCCCTCTTGGGCTTCC 121
QY 2126 CAGACACTAAGTCTGGAATGAAATTCACCTGCCTCTGAATTCGCCACCTGGTGGGGCGAG 2185
DB 122 CAGACACTAAGTCTGGAATGAAATTCACCTGCCTCTGAATTCGCCACCTGGTGGGGCGAG 181
QY 2186 GGGTGTGACTTGGCTTCCAGCTCGAAGATATCTACCCAGCCCTAGCTATATAACGG 2245
DB 182 GGGTGTGACTTGGCTTCCAGCTCGAAGATATCTACCCAGCCCTAGCTATATAACGG 239
QY 2246 GCTGTGTGGAGGGCTCCACAGGGCCAGTTCACAGGGTTCATCCACAGAGAGAGAAAC 2305
DB 240 GCTGTGTGGAGGGCTCCACAGGGCCAGTTCACAGGGTTCATCCACAGAGAGAGAAAC 298
QY 2306 ATAGACTCGAGG 2317
DB 299 ATAGACTCAGG 310

RESULT 5
BI293043/c 478 bp mRNA linear EST 19-JUL-2001
LOCUS UI-R-DKO-cdj-f-07-0-UI.s1 UI-R-DKO Rattus norvegicus cDNA clone
DEFINITION UI-R-DKO-cdj-f-07-0-UI 3', mRNA sequence.
ACCESSION BI293043
VERSION BI293043.1 GI:14954159
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Rattus.
REFERENCE 1 (bases 1 to 478)
AUTHORS Ronaldo, M.F., Lennon, G. and Soares, M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: benito-soares@uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to identify it as a clone from the

```

normalized rat heart pool library cDNA Library Preparation: M.B. Soares Lab Clone distribution: clones will be available through Research Genetics (www.resgen.com) The following repetitive elements were found in this cDNA sequence: 170-292, >URR1A#DNA/MER1 type
Seq primer: M13_Foward
POLYA=yes:

FEATURES

Location/Qualifiers

1. .478
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-DKO-cdj-f-07-0-UI"
/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-DKO"
/note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site 1: Not 1; Site 2: Eco RI; The UI-R-DKO library is a subtracted library derived from a mixture of five individually tagged normalized rat libraries: brain-nRBP (20%), heart-nRHP (20%), kidney-nRKP (20%), aorta-nRAP (20%), and placenta-nRPP (20%). Each original library was constructed from a mixture of equal amounts of RNA from seven different developmental time-points: embryonic day 17, embryonic day 19, embryonic day 21, adult day 1, adult day 12, adult day 75, and adult day 200. (Exception: the aorta pool does not contain embryonic day 17 RNA and the placenta pool contains only the three embryonic stages). Each library was normalized individually according to the procedure described by Bonaldo, Lennon & Soares (Genome Research 6: 791-806, 1996). For construction of the DKO subtracted library, plasmid DNA from each of the five individually tagged normalized libraries was mixed in the proportions specified above and electroporated into competent bacteria for production of single-stranded circular DNA representing the pool of libraries. Single-stranded circular DNA representing these five normalized libraries was then used as a tracer in a subtractive hybridization with a driver (PCR amplified inserts from a plasmid DNA template preparation) comprising: a) a set of about 1,000 arrayed clones from each of the five non-normalized libraries of brain (C70s), heart (C50s), kidney (C40s), aorta (C60s), and placenta (C50s). The resulting pool of approximately 5,000 clones represented about 33.3% of the final driver population. A set of about 2,000 arrayed clones from each of the five normalized libraries of brain (C70), heart (C50), kidney (C40), aorta (C60), and placenta (C50). The resulting pool of about 10,000 clones represented about 66.6% of the final driver population.
TAG_LIB=UI-R-DKO
TAG_TISSUE=rat heart pool
TAG_SEQ=ATAAGATAC"

BASE COUNT 118 a 103 c 110 g 146 t 1 others

Query Match 8.1%; Score 191.6; DB 12; Length 478;
Best Local Similarity 69.5%; Pred. No. 7.4e-41;
Matches 372; Conservative 0; Mismatches 80; Indels 83; Gaps 5;

QY 1 GGATCTTTTCATGTTTAAACATATCAACCTTAACCCAGGGGAAACAGCTGCTGACAGT 60
|||||
DB 475 GGATCTCTTAATGTTTAAACATATCAACAGTAGCCCA--GGGAAAGACTGCTGACAGT 418
QY 61 GCGTTTGCCACCATGAATACCTCTAGTGTAGTCCGTTTGTGAAGAACTCAGCCCATCCCA 120
|||||
DB 417 GCGTNTGCCACCATGAACACATCTAGTCTCCCTCCCTTTGTGAATTCAGCCCATCCCA 358
QY 121 ACATCTCTGCAAGCCCATCTCTACAGGTGCTCATTTGGGAATTTCTGGAGCTTCTCT 180
|||||
DB 357 ACA-----GTTCTGTCGCCGAGCTTCTCT 333

QY 181 TTCAGATCAGCTGATTTAGGGCAGCAGTCTTCAACCTGGGGCTCGACCCCTTTGG 240
DB 332 TTCAGGGTCAGCTGATTTAGGGCAGCAGTCTTCAACCTGGA----- 290
QY 241 GGGATCAACAGCCTTTACAGGGGTACATATCATCTCTATATGTCAGGTATTTA 300
DB 289 -----GTCATATCATCTCTATCTATATCAAAATATTTCA 255
QY 301 CATTACGATTCGTAAACAGTAGCAAAATTACAGGTATGAAATAGCAATGAATATTTAT 360
DB 254 CATCATGACTCATACAGTAACACATTTACAGGCATGATGACATGAATGATTTAT 195
QY 361 GATTGAAGTCAACCAACATGAGGCCGACACTCTTCTAGAGAAAAATCACCTGGTG 420
DB 194 GGTGAGGGTCAACACCATGAGGCCACACACTGTTCTGGGGAA---CAGTGGAGAG 138
QY 421 GGGAAAGGTTTGGGAAGCCTTCTGTCCATCTTCTTCAATCTTCAAGTGTGTTTCACA 480
DB 137 AGGAAAGGATTGAGGAAGCC-TTCCATCTGTTCTGCACCTCTTCAAGGGATGTATTACA 79
QY 481 GAAAGCCTTTTCAGCTGTTCTGCTGGGGCTCTTAGTAAGTCTGAGTAGGAACCTGTA 535
DB 78 GAAAGCCTTTGCTGGGTTTGTGTTGGGCTCTTAGTAATCTGAGCAGAAACTGCA 24

RESULT 6

BF543101

LOCUS

DEFINITION

UI-R-AG1-aal-b-08-0-UI.r1 UI-R-AG1 Rattus norvegicus cDNA clone

BF543101

VERSION

BF543101.1 GI:11634214

KEYWORDS

SOURCE

ORGANISM

Rattus norvegicus (Norway rat)

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;

Rattus.

1 (bases 1 to 434)

Bonaldo, M.F., Lennon, G. and Soares, M.B.

Normalization and subtraction: two approaches to facilitate gene

discovery

Genome Res. 6 (9), 791-806 (1996)

97044477

8889548

Contact: Soares, MB

Coordinated Laboratory for Computational Genomics

University of Iowa

375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA

Tel: 319 335 8250

Fax: 319 335 9565

Email: bent-soares@uiowa.edu

cDNA Library Preparation: M.B. Soares Lab Clone distribution:

clones will be available through Research Genetics (www.resgen.com)

This clone is also available through the I.M.A.G.E. Consortium at

LLNL (info@image.llnl.gov). IMAGE ID= 1790595 The following

repetitive elements were found in this cDNA sequence: 45-167,

>URR1A#DNA/MER1 type

Seq primer: M13_Foward.

Location/Qualifiers

1. .434

/organism="Rattus norvegicus"

/mol_type="mRNA"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-AG1-aal-b-08-0-UI"

/dev_stage="adult"

/lab_host="DH10B (Life Technologies)"

/clone_lib="UI-R-AG1"

/note="Vector: p773D-Pac (Pharmacia) with a modified

polylinker; Site 1: Not 1; Site 2: Eco RI; The UI-R-AG1

library is a normalized library constructed from 13 dpc

rat ventricle. The tag is a string of 6 nucleotides

present between the Not I site and the oligo-dT track.
The library was constructed as described by Bonaldo,
Lennon and Soares, Genome Research 6: 791-806, 1996.
Tissue provided by Jim Lin, Department of Biology,
University of Iowa.

```

BASE COUNT      119 a   94 c   103 g   117 t   1 others
ORIGIN

Query Match      7.8%; Score 183; DB 10; Length 434;
Best Local Similarity 73.0%; Pred. No. 1.6e-38;
Matches 276; Conservative 0; Mismatches 96; Indels 6; Gaps 3;

QY 261 CAGGGTCACATATCATCTATCTATATGTCAGGTATTTACATTACGATTGTAACAGTA 320
Db 43 CTGGAGTCACATATCATCTATCTGTCATATCAATATTCACATCATGACTCATAACAGTA 102
QY 321 GCAAAATTACAGTATGAATACGATGAATATTTATGATGAGGTACCAACA 380
Db 103 ACAAAATTACAGCATGATGATGATGAATATTTATGATGAGGTACCAACA 162
QY 381 TGAGGGCCGACACTGTTCTAGAGAAAATCACCTGGTGGGAAAGGTTTGGGAAGCC 440
Db 163 TGAGGGCCGACACTGTTCTGGGGAA---CATGGAGAGAGAAAGGATTGAGGAAGCC 219
QY 441 TTCTCTCCATCTTCTCAATCTTCAAGTGATGTGTTACAGAAAGGCTTTTCAGCTGTCT 500
Db 220 -TTCCATCTGTTCTGCACCTTCAAGGGATGATTCACAGAAAGGCTTTGTCGTTT 278
QY 501 GCTGGGGCTTTAGTAGTCTAGTAGGAACTGATGATGATGATGATGATGATGATGATG 560
Db 279 GTTGGGGCTTTAGTAGTCTAGTAGGAACTGATGATGATGATGATGATGATGATGATG 336
QY 561 GGAGCCAAAGCAGCATCGTGGTGGAGGAGAGCAACCTTCTAGCTCTGATCC 620
Db 337 TGATTTTCTGCTTCTCTGNGTGAACCAAGACTGGACTCCATGTCAGCCCTGCATCT 396
QY 621 ATAGCAAGTAGCCTAATG 638
Db 397 GCAGCTAGTTGCCTAATG 414

```

```

RESULT 7
AW252440/c
LOCUS
DEFINITION
UI-R-BJ0-adx-e-05-0-UI.s1 UI-R-BJ0 Rattus norvegicus cdna clone
UI-R-BJ0-adx-e-05-0-UI 3', mRNA sequence.
ACCESSION
AW252440
VERSION
AW252440.1 GI:6596031
SOURCE
Rattus norvegicus (Norway rat)
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 488)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
JOURNAL
MEDLINE
PUBMED
COMMENT
Contact: Soares, MB
Coordinated Laboratory for Computational Genomics
University of Iowa
375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: bento-soares@uiowa.edu
The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cdna and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cdna between the NotI site
and the oligo-dT track served to identify it as a clone from the

```

normalized AV canal at 15 dpc library cdna Library Preparation:
M.B. Soares Lab Clone Distribution: Clones will be available
through Research Genetics (www.resgen.com) The following repetitive
elements were found in this cdna sequence: 1-44,
>POLY_ASimple_repeat
Seq primer: M13 Forward
POLYA=Yes.

```

FEATURES
source
Location/Qualifiers
1..488
/organism="Rattus norvegicus"
/mol_type="mRNA"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
/clone="UI-R-BJ0-adx-e-05-0-UI"
/dev_stage="adult"
/lab_host="DH10B (Life Technologies)"
/clone_lib="UI-R-BJ0"
/notes="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BJ0
library is a subtracted library derived from the UI-R-AA1,
UI-R-AB1, UI-R-AC1, UI-R-AD1, UI-R-AE1, UI-R-AF1, and
UI-R-AG1 libraries. These libraries represent tissues from
rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal
at 16.5 dpc, atrium at 15 dpc, ventricle at 13 dpc, AV
canal at 15 dpc, and ventricle at 13 dpc. The tag is a
string of 5-6 nucleotides present between the Not I site
and the oligo-dT track. The library was constructed as
described by Bonaldo, Lennon and Soares, Genome Research
6: 791-806, 1996.
TAG LIB=UI-R-BJ0
TAG TISSUE=AV canal at 15 dpc
TAG_SEQ=GAAGG"
BASE COUNT      115 a   111 c   106 g   155 t   1 others
ORIGIN

```

```

Query Match      7.3%; Score 171; DB 9; Length 488;
Best Local Similarity 78.4%; Pred. No. 3.3e-35;
Matches 261; Conservative 0; Mismatches 55; Indels 17; Gaps 4;

QY 879 ATGAAACAGAGAACCAACAGTATTCATTCAGCTCTCAGGACAGATGACAGAG 938
Db 349 ATGGAACAGAGAACCAACAGTATTCATTCAGCTCTCAGGACAGATGACAGAG 290
QY 939 AGAACACTAGGAGAGGGAGAACCAACAGTATTCAGCTCTCAGGACAGATGACAGAG 998
Db 289 -----CACTAGGAGAGAGAACCAACAGTATTCAGCTCTCAGGACAGATGACAGAG 239
QY 999 AATGCTCTTCTACTGAAGATTCTAGAAACACAAATTTCTGTTGAACAGCTGAAGTGGGT 1058
Db 238 AATGCTCTTCTACTGAAGATTCTAGAAACACAAATTTCTGTTGAACAGCTGAAGTGGGT 179
QY 1059 GGGG---GTTCTTACCCCATGTTCTATGGAAGGTTAGTGAAGAGACAGATATATGATG 1115
Db 178 GGGGTGAAGATAACCAACCACTGTTCTAGGAGGTTAGGAGGAGACAGATATATGATG 119
QY 1116 GCAGACATACAAACATACACACACCTTAATTAACACTTCCCTCTTCTACTGACACCCC 1175
Db 118 GCAGCAT-----ACAAACATACACCTTAATTAATGCTTCCCTCTGCTACTGACACTCT 64
QY 1176 CTTCACTCTCTCTCTTTCATAAAAAATAAAAAA 1208
Db 63 CCCTTCCTCTCTCTTTCATAAAAAATAAAAAA 31

RESULT 8
AW251306/c
LOCUS
DEFINITION
UI-R-BJ0-adx-f-04-0-UI.s1 UI-R-BJ0 Rattus norvegicus cdna clone
UI-R-BJ0-adx-f-04-0-UI 3', mRNA sequence.
ACCESSION
AW251306
VERSION
AW251306.1 GI:6594897
SOURCE
Rattus norvegicus (Norway rat)

```

```

ORGANISM      Rattus norvegicus
               Mammalia; Eutheria; Chordata; Craniata; Vertebrata; Euteleostomi;
               Murinae; Eukaryota; Metazoa;
REFERENCE     1 (bases 1 to 511)
AUTHORS       Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE         Normalization and subtraction: two approaches to facilitate gene
               discovery
JOURNAL       Genome Res. 6 (9), 791-806 (1996)
MEDLINE       97044477
PUBMED       8889548
COMMENT       Contact: Soares, MB
               Coordinated Laboratory for Computational Genomics
               University of Iowa
               375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
               Tel: 319 335 8250
               Fax: 319 335 9565
               Email: bento-soares@uiowa.edu
               The sequence contained an oligo-dT track that was present in the
               oligonucleotide that was used to prime the synthesis of first
               strand cDNA and therefore this may represent a bonafide poly A
               tail. The sequence tag present in the cDNA between the NotI site
               and the oligo-dT track served to identify it as a clone from the
               normalized ventricle at 13 dpc library cDNA library Preparation:
               M.B. Soares Lab Clone distribution: clones will be available
               through Research Genetics (www.resgen.com)
               Seq primer: M13 Forward
               POLYA=Yes.

FEATURES             Location/Qualifiers
     source           1..511
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-BJ0-adg-f-04-0-UI"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="UI-R-BJ0"
                     /note="Vector: pT73D-Pac (Pharmacia) with a modified
                     polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-BJ0
                     library is a subtracted library derived from the UI-R-AAL,
                     UI-R-ABI, UI-R-ACL, UI-R-ADI, UI-R-AEI, UI-R-AFI, and
                     UI-R-AGI libraries. These libraries represent tissues from
                     rat atrium at 16.5 dpc, ventricle at 16.5 dpc, AV canal
                     at 16.5 dpc, atrium at 15 dpc, ventricle at 15 dpc, AV
                     canal at 15 dpc, and ventricle at 13 dpc. The tag is a
                     string of 5-6 nucleotides present between the Not I site
                     and the oligo-dT track. The library was constructed as
                     described by Bonaldo, Lennon and Soares, . Genome Research
                     6: 791-806, 1996.
                     TAG_LIB=UI-R-BJ0
                     TAG_TISSUE=ventricle at 13 dpc
                     TAG_SEQ=CAGCGA"

BASE COUNT      127 a 120 c 113 g 151 t
ORIGIN
Query Match      7.3%; Score 171; DB 9; Length 511;
Best Local Similarity 78.4%; Pred. No. 3.3e-35;
Matches 261; Conservative 0; Mismatch 55; Indels 17; Gaps 4;

QY      879  ATGAACAGAGAGACCAACAGTTATCCATTGATAGCGTCTCAGGACAGATAGGACAGAG 938
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      331  ATGGAACAGAGAGGCCCAACAGTTGTCCTTCGATAGTGCTCAGGACAGCCAGGACAGAG 272
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      939  AGAACACTAGGAGAGGGGACCCAGCAAGGACCAAGGATATAGTGTGTTGGTTTTCAGGGC 998
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      271  ----CACTAGGAGAGGAGAACCCAGCAAGGA-----TATCAGTGTGCTGTTTCCAGGCG 221
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      999  AATGTCTTCTACTAGATTCTTAGAACACAATTTGCTGTTTCACAGCTGAAGTGGGCT 1058
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      220  AATGTCTTCTCCGAAGGTTCTAGAACACAGATTTACTGGATGAGAGCTGAAGTGGGCT 161
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      1059 GGGG---GTTCTTACCCCTGTTTCATGAAGAGGCTGAGTGAGGAGAGACAGATATATGATG 1115
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

Db      160  GGGGTGAAGATAACACCACTTCATCTGAAGGCTGAGGAGGAGCGACAGACATACGATG 101
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      1116 GCACGATAACAAACATACACACACCCCTAAATTAACACACTTCCTCTTCTACTGACACCCC 1175
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      100  GCACGAT-----ACAACATACACACCCCTAAATTAATGCTTCCTCTGCTACTGACACTCT 46
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY      1176 CTTCACTCTCTCTCTTTTCATATAAAAAATAAAAAA 1208
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      45  CCCTTCACCTCTCTTTTCATATAAAAAATAAAAAA 13
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 9
LOCUS      AI710529/c
DEFINITION UI-R-AG1-aal-b-08-0-UI-s1 UI-R-AG1 Rattus norvegicus cDNA clone
            UI-R-AG1-aal-b-08-0-UI 3', mRNA sequence.
ACCESSION  AI710529
VERSION     AI710529.1 GI:5000305
KEYWORDS   EST.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE  1 (bases 1 to 370)
AUTHORS    Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE      Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL    Genome Res. 6 (9), 791-806 (1996)
MEDLINE    97044477
PUBMED     8889548
COMMENT    Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to verify it as a clone from the
            normalized ventricle at 13 dpc library cDNA Library Preparation:
            M.B. Soares Lab Clone distribution: clones will be available
            through Research Genetics (www.resgen.com) The following repetitive
            elements were found in this cDNA sequence: 1-43,
            >POLY_A#Simple_repeat
            Seq primer: M13 Forward
            POLYA=Yes.

FEATURES             Location/Qualifiers
     source           1..370
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-AG1-aal-b-08-0-UI"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="UI-R-AG1"
                     /note="Vector: pT73D-Pac (Pharmacia) with a modified
                     polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-AG1
                     library is a normalized library constructed from 13 dpc
                     rat ventricle. The tag is a string of 6 nucleotides
                     present between the Not I site and the oligo-dT track.
                     The library was constructed as described by Bonaldo,
                     Lennon and Soares, . Genome Research 6: 791-806, 1996.
                     Tissue provided by Jim Lin, Department of Biology,
                     University of Iowa.
                     TAG_LIB=UI-R-AG1
                     TAG_TISSUE=ventricle at 13 dpc
                     TAG_SEQ=CAGCGA"

```



```

/sex="Male"
/cell_type="Lymphocytes"
/clone_lib="RPCI-11"
/note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
RPC11 Human Male BAC Library"
BASE COUNT      168 a   100 c   113 g   141 t
ORIGIN

Query Match      4.9%; Score 116; DB 28; Length 522;
Best Local Similarity 76.7%; Pred. No. 3.9e-20;
Matches 155; Conservative 0; Mismatches 45; Indels 2; Gaps 1;

QY 1540 TCTCTCTGTGCATCACTCGCGCGCTTTGGGT--AGATCTCTCTAGCTTCAGAT 1597
Db 202 TCTATTTCTTGACCATCTGATCCATTTTGAAGTAAATGCTCCNAATTATTATGCTGT 143
QY 1598 TTAGAACCGGTGAGCTGTGTGCTACTAATTATGGCCAGTGACACCATAGAGTCAAAGT 1657
Db 142 TTAGAACCGGTAAAGCATGTGTGCTAATTATGGCCAGTGACATCATAAAGAAAGT 83
QY 1658 GCATTAAGTAAAGCTTTCAATTTCTCTAATGCTGTGATGGATGTCACAGGGCCAT 1717
Db 82 GCATTAAGTAAAGCTTTCAATTTCTTATAATGATGTAAGTGGCATGTGTCATGGGCGCTA 23
QY 1718 TTTAGCTGCAGACATCACTCCA 1739
Db 22 TTTAGCCCGACATCACTCCA 1

RESULT 14
LOCUS AZ114168 518 bp DNA linear GSS 12-MAY-2000
DEFINITION RPCI-23-449H23-TJ RPCI-23 Mus musculus genomic clone RPCI-23-449H23
, genomic survey sequence.
ACCESSION AZ114168
VERSION AZ114168.1 GI:7774139
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 518)
AUTHORS Zhao, S., Niernan, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret
, B., Levins, M., Megann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished
COMMENT Other GSSs: RPCI-23-449H23-TV
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pietere@jeng.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 449 row: H column: 23
Seq primer: SP6
Class: BAC ends.
FEATURES
source
1. .518
/organism="Mus musculus"
/mol_type="genomic DNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-449H23"
/sex="Female"
/lab_host="DH10B"

/clone_lib="RPCI-23"
/note="Organ: Kidney/Brain; Vector: pBACe3.6; Site 1:
EcoRI; Site 2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACe3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT      158 a   94 c   108 g   158 t
ORIGIN

Query Match      4.9%; Score 114.6; DB 28; Length 518;
Best Local Similarity 67.2%; Pred. No. 9.4e-20;
Matches 162; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 145 ACAAGGTGCTCATTTGGGAATTTCTGGAGCTTCTTTTCAGGATCAGCTGATTCTAGGG 204
Db 150 ACAGGCCATGCAATTTGAAAGCTTAGTGTCTTCACTTTTAACTACTTTTGACGACAG 209
QY 205 CAGCAGTTCTCAACCTGGGGGCTCGACCCCTTTGGGGAAATCAACGACCCCTTTACAGG 264
Db 210 CAGTGTCTCAACCTGGGTGGGACCCCTGTGGGGATTACTGTCTCTCCAGG 269
QY 265 GGTCAATATCATCTATCTATATGTTCAGGTATTTACATTAGCATTCGTAACAGTAGCAA 324
Db 270 GGTCTCTTATCAATATTTCCACAAGTCAGATATGTATATTACAGTTTATAACAATGGCAA 329
QY 325 AATTACAGGTATGAATAGCAATGAATATTTTATGATTGAAGTCACACACATGAG 384
Db 330 AATTCCAGTTATGAGGAGTAATGAATATTTTATGTTAGTGGGTTCATCACAACATGAG 389
QY 385 G 385
Db 390 G 390

RESULT 15
LOCUS AZ290439/c 596 bp DNA linear GSS 27-JUL-2000
DEFINITION RPCI-23-59022.TVB RPCI-23 Mus musculus genomic clone RPCI-23-59022,
genomic survey sequence.
ACCESSION AZ290439
VERSION AZ290439.1 GI:9532225
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Mus musculus
REFERENCE 1 (bases 1 to 596)
AUTHORS Zhao, S., Niernan, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret
, B., Levins, M., Megann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
and Fraser, C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished
COMMENT Other GSSs: RPCI-23-59022.TJB
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pietere@jeng.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (http://bacpac.med.buffalo.edu/orderingframe.htm)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tldb/bac_ends/mouse/bac_end_intro.html
Plate: 59 row: G column: 22
Seq primer: T7
Class: BAC ends.
FEATURES
source
1. .596
Location/Qualifiers

```


THIS PAGE BLANK (USPTO)